

(11)特許出願公開番号

特開平9-101990

(43)公開日 平成9年(1997)4月15日

(51) Int.Cl.⁸

G O 6 F 17/60
17/30

識別記号

室内整理番号

FI

G O 6 F 15/21
15/40

$$\mathbf{z}$$

370A

技術表示箇所

審査請求 未請求 請求項の数6 FD (全 57 頁)

(21)出願番号	特願平7-335790
(22)出願日	平成7年(1995)11月30日
(31)優先権主張番号	特願平7-212939
(32)優先日	平7(1995)7月31日
(33)優先権主張国	日本(JP)

(71)出願人	000003078	
	株式会社東芝	
	神奈川県川崎市幸区堀川町72番地	
(72)発明者	酒井 哲也	
	神奈川県川崎市幸区小向東芝町 1 番地	株
	式会社東芝研究開発センター内	
(72)発明者	三池 誠司	
	神奈川県川崎市幸区小向東芝町 1 番地	株
	式会社東芝研究開発センター内	
(72)発明者	住田 一男	
	神奈川県川崎市幸区小向東芝町 1 番地	株
	式会社東芝研究開発センター内	
(74)代理人	弁理士 鈴江 武彦	

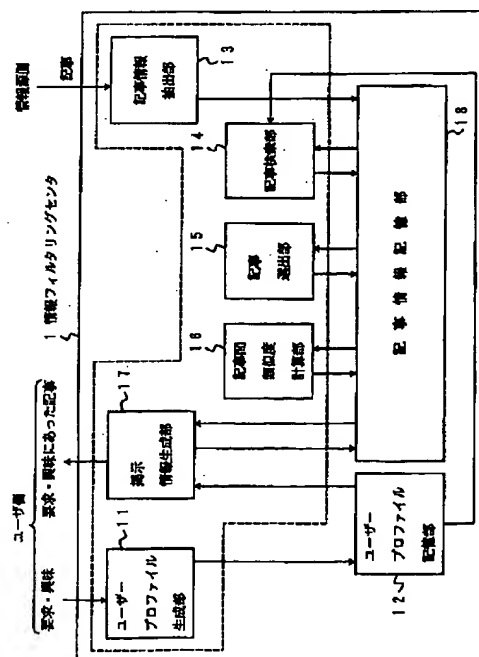
[最終頁に続く](#)

(54) 【発明の名称】 情報フィルタリング装置

(57) 【要約】

【課題】 情報フィルタリングによって選択された記事に関連記事の情報を付加してユーザに送信し、送信記事の有効利用を図る。

【解決手段】情報フィルタリングセンタ１内には記事間の類似度を算出する記事間類似度計算部１６が設けられており、この記事間類似度計算部１６によって重複記事の存在が調べられる。重複記事を構成する記事群の１つがユーザに提示する記事として選択され、他の記事は排除される。この場合、排除された記事に関する情報は、関連記事情報として選択記事に付加されてユーザに送られる。よって、ユーザに提供される記事同士の関連性をユーザに提示できるようになり、ユーザによる送信記事の有効利用を図る事が可能になる。



(2)

2

【特許請求の範囲】

【請求項1】 複数の情報源からテキストやイメージなどの記事の配信を受け、それら配信された記事の中から所定の記事を選出してユーザに提示する情報フィルタリング装置において、

ユーザ毎に予め指定された検索条件を保持する手段と、配信された記事を検索し、ユーザ毎に検索条件に合致する記事を選定する記事検索手段と、

この記事検索手段によって選定された記事同志または選定された記事と他の記事との間の類似度を算出し、その類似度に従って記事毎に関連記事を選定する手段と、決定された関連記事の情報を前記選定された記事に付加してユーザに提示する手段とを具備することを特徴とする情報フィルタリング装置。

【請求項2】 複数の情報源からテキストやイメージなどの記事の配信を受け、それら配信された記事の中から所定の記事を選出してユーザに提示する情報フィルタリング装置において、

ユーザ毎に予め指定された検索条件を保持する手段と、配信された記事を検索し、ユーザ毎に検索条件に合致する記事を選定してユーザに提示する記事検索手段と、この記事検索手段によって選定された記事が満足した検索条件を示す情報を各記事に付加してユーザに提示する手段を具備することを特徴とする情報フィルタリング装置。

【請求項3】 複数の情報源からテキストやイメージなどの記事の配信を受け、それら配信された記事の中から所定の記事を選出してユーザに提示する情報フィルタリング装置において、

ユーザ毎に予め指定された検索条件を保持する手段と、配信された記事を検索し、ユーザ毎に検索条件に合致する記事を選定してユーザに提示する記事検索手段と、この記事検索手段によって選定された記事の種類に応じた長さの要約または抄録を生成し、その要約または抄録をユーザに提示する手段を具備することを特徴とする情報フィルタリング装置。

【請求項4】 少なくとも1つ以上の情報源からテキストやイメージなどの記事の配信を受ける手段と、あらかじめユーザが指定した検索条件と配信された記事との類似度を算出する手段と、算出した類似度の順に記事をソートして、一定の数の記事、あるいはあらかじめ定めた閾値以上の類似度を有する記事のみを類似度の順で出力する出力手段を有する情報フィルタリング装置において、

記事間の類似度を算出する手段を具備し、その算出した記事間類似度にしたがって記事のグループ化、関連づけ、あるいは出力記事の選択制御を行うことを特徴とする情報フィルタリング装置。

【請求項5】 前記記事間の類似度を算出する手段は、一文目、一段落目、見出しといった書式上のフィールド

ごとに記事間で類似度を求め、それらの荷重平均を記事間の類似度とすることを特徴とする請求項4記載の情報フィルタリング装置。

【請求項6】 少なくとも1つ以上の情報源からテキストやイメージなどの記事の配信を毎日定期的に受ける手段と、あらかじめユーザが指定した検索条件と配信された記事との類似度を算出する手段と、算出した類似度の順に記事をソートして、一定の数の記事、あるいはあらかじめ定めた閾値以上の類似度を有する記事のみを選択する情報フィルタリング装置において、

フィルタリング結果としてユーザに出力した記事を記憶する出力記事記憶手段と、

この出力記事記憶手段に記憶されている記事と当日配信された記事とを合わせてそれら記事間で類似度を算出し、その類似度にしたがって記事のグループ化あるいは関連づけを行ってユーザに出力する手段とを具備し、当日の記事だけからなる記事グループか、それ以前の日の記事も含まれているのかを区別するための情報を出力記事に付加することを特徴とする情報フィルタリング装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、膨大なテキスト記事からユーザの要求・興味にあったものを選出して定期的にユーザに提供する情報フィルタリング装置に関する。

【0002】

【従来の技術】 近年、ワードプロセッサや電子計算機の普及、および計算機ネットワークを介した電子メールや電子ニュースの普及などに伴い、文書の電子化は加速的に進みつつある。

【0003】 電子出版という言葉が示すように、今後は新聞、雑誌、本の情報も電子的に提供されることが一般的になると考えられる。これにより、個人にとってリアルタイムで入手可能となるテキスト情報の量は膨大になっていくと予測される。

【0004】 これに伴い、新聞や雑誌などの膨大なテキスト記事からユーザの要求・興味にあったものを選出して定期的にユーザに提供する情報フィルタリングシステムあるいは情報フィルタリングサービスの需要が高まりつつある。

【0005】 従来より実現されている情報フィルタリングシステムは、ユーザの要求や興味を表現したユーザプロフィールに合致する記事を検索し、これらの見出しのリストあるいは記事全体をならべてユーザに提示するものである。

【0006】 通常、ユーザプロフィールは、ユーザが興味をもっているトピックをいくつか指定することにより作成される。

【0007】 また、提示された記事の有用性の判定をユ

ーザが行い、この情報をユーザプロフィールに反映させることにより、次回以降の情報フィルタリングの適合率を高めるといったレバンス・フィードバックという機能が実現されている。

【0008】

【発明が解決しようとする課題】しかし、従来のシステムでは、選定された記事が羅列されてユーザに提示されるだけであるので、今回提示された記事同士の関係や、今回提示された記事と前回までに提示された記事との関係を把握することがユーザにとっては難しいという問題点があった。

【0009】また、従来のような記事の単純な提示では、いかなるトピックのいかなる検索条件にマッチしたためにその記事がユーザに提示されるに至ったか、また、提示された記事を他のユーザはどのように読んでいるかなどの情報が欠落していたため、有用性の判定には多大な労力を必要とし、その一貫性を保つことが難しいという問題点があった。

【0010】また、情報フィルタリングシステムにおいては、重要記事を選出した後にさらに、その記事内の重要なテキストを部分的に選出するという二重のフィルタリングを行うことは、長い記事からの情報収集を効率的にするという点で有効である。しかし、従来では、機械的に適当な長さのテキストを抜粋表示しているにすぎなかったため、余分な情報が混在していたり、必要な情報が欠けていたりするという問題点があった。

【0011】また、従来では、ニュースソースから配信されるテキストと検索条件との類似度にしたがってユーザに提供するテキストを選択するだけであったため、同じ内容のテキストであってもばらばらに出力される等の問題があった。

【0012】この発明は上述の事情に鑑みてなされたものであり、情報フィルタリングによってユーザに提供される記事同士の関連性をユーザに提示できるようにし、ユーザに記事同士の関連性を把握させることが可能な情報フィルタリング装置を提供することを第1の目的とする。

【0013】また、この発明は、提示されている記事がいかなる検索条件を満足したのかがユーザにわかるようにすることにより、情報フィルタリングに対するユーザの理解と信頼を深めることが可能な情報フィルタリング装置を提供することを第2の目的とする。

【0014】さらに、この発明は、記事の種類に応じてユーザに提示する要約あるいは抄録の長さを調節できるようにし、2重フィルタリングを効率的に行うことができる情報フィルタリング装置を提供することを第3の目的とする。

【0015】また、この発明は、互いに内容が重複する記事同士をグループ化あるいは関連づけてユーザに提供できるようにし、ユーザがテキスト記事を読むための

手間を大幅に軽減することが可能な情報フィルタリング装置を提供することを第4の目的とする。

【0016】

【課題を解決するための手段】この発明は、複数の情報源からテキストやイメージなどの記事の配信を受け、それら配信された記事の中から所定の記事を選出してユーザに提示する情報フィルタリング装置において、ユーザ毎に予め指定された検索条件を保持する手段と、配信された記事を検索し、ユーザ毎に検索条件に合致する記事を選定する記事検索手段と、この記事検索手段によって選定された記事同志または選定された記事と他の記事との間の類似度を算出し、その類似度に従って記事毎に関連記事を決する手段と、決定された関連記事の情報を前記選定された記事に付加してユーザに提示する手段とを具備することを特徴とする。

【0017】この情報フィルタリング装置においては、例えば記事表現を記事間で比較することによって記事同志の類似度が算出され、その類似度に従って、ユーザに提示される記事とそれに関連する関連記事が決定される。この関連記事の情報は、ユーザに提示される記事の本文情報などに付加されてユーザに送られる。類似度計算を行う対象としては、今回到着した記事同士、または今回到着した記事と前回までに到着した記事との間で行うことが好ましい。これにより、記事検索手段によって選定されている記事同士の関係や、今回選定された記事と過去のフィルタリングで選定された記事との関係などが明確になり、記事同士の関連性などをユーザに知らせることができる。

【0018】また、記事検索手段によって選定された記事同志の類似度を算出することによって重複記事の存在を調べれば、重複記事の本文情報についてはユーザに提示せず、その重複記事の見出しなどの情報だけを関連記事情報として付加してユーザに提示することもできる。これにより、例えば異なる複数の情報源から得られた同一内容に関する記事が、重複してユーザに提示されることを自動的に回避することができる。

【0019】また、この発明は、複数の情報源からテキストやイメージなどの記事の配信を受け、それら配信された記事の中から所定の記事を選出してユーザに提示する情報フィルタリング装置において、ユーザ毎に予め指定された検索条件を保持する手段と、配信された記事を検索し、ユーザ毎に検索条件に合致する記事を選定してユーザに提示する記事検索手段と、この記事検索手段によって選定された記事が満足した検索条件を示す情報を各記事に付加してユーザに提示する手段を具備し、その記事が選択された根拠をユーザに知らせることができるようにしたことを特徴とする。

【0020】この構成によれば、提示されている記事がユーザの選択したトピックのうちいずれに適合したものであるかなど、提示されている記事がいかなる検索条件

を満足したのかがユーザに明示されるため、何故その記事が提示されているかがユーザに分かりやすくなり、記事の有用性の判定が容易になる。

【0021】したがって、既にユーザに送信された各記事がユーザにとって有用であったか否かなどの情報についてユーザからフィードバックを受け、その情報を反映して検索条件を修正するレlevance・フィードバック機能をさらに設けることにより、記事が選択された根拠の提示を、そのレlevance・フィードバック機能に有効利用できるようにする。

【0022】また、記事が選択された根拠の代わりに、提示されている記事が他のユーザによってどのように読まれているかをユーザに提示することによっても、他のユーザの判断を参考にしたレlevance・フィードバックが可能になり、レlevance・フィードバックの有効利用を図ることができる。

【0023】また、この発明は、複数の情報源からテキストやイメージなどの記事の配信を受け、それら配信された記事の中から所定の記事を選出してユーザに提示する情報フィルタリング装置において、ユーザ毎に予め指定された検索条件を保持する手段と、配信された記事を検索し、ユーザ毎に検索条件に合致する記事を選定してユーザに提示する記事検索手段と、この記事検索手段によって選定された記事の種類に応じた長さの要約または抄録を生成し、その要約または抄録をユーザに提示する手段を具備することを特徴とする。

【0024】この構成によれば、記事の種類に応じた長さの要約あるいは抄録が作成されてそれがユーザに提示されるため、ユーザに提示されるテキストのうち、ユーザにとって有用であるテキスト情報の占める割合が高くなる。これにより、効率的な情報収集が可能になる。

【0025】記事の種類区分としては、その記事が満足したトピックなどの検索条件の違いや、その記事の発行日時などの記事そのものの属性の違いなどを利用することが好ましい。例えば、ユーザが検索条件として複数のトピックを指定し、且つそれらトピックに優先度を設定した場合には、優先度の高いトピックに該当する記事が検索される程、要約・抄録のサイズを長くでき、ユーザにとって有用であるテキスト情報の占める割合が高くなる。

【0026】また、この発明は、少なくとも1つ以上の情報源からテキストやイメージなどの記事の配信を受ける手段と、あらかじめユーザが指定した検索条件と配信された記事との類似度を算出する手段と、算出した類似度の順に記事をソートして、一定の数の記事、あるいはあらかじめ定めた閾値以上の類似度を有する記事のみを類似度の順で出力する出力手段を有する情報フィルタリング装置において、記事間の類似度を算出する手段を具備し、その算出した記事間類似度にしたがって記事のグループ化、関連づけ、あるいは出力記事の選択制御を行

うことを特徴とする。

【0027】この情報フィルタリング装置においては、関連する記事同士がグループ化あるいは関連づけて利用者に提供することが可能となる。従来のように関連のあるテキストが、順不同で出力された場合には、利用者はテキストごとに内容を理解するために頭を切り替える必要が生じ、フィルタリング結果全体を理解するための手間がかかることになるが、この発明の情報フィルタリング装置では、関連する記事同士がグループ化あるいは関連づけて利用者に提供されるため、利用者の手間を大幅に軽減することができる。

【0028】記事間類似度は当日配信された記事間のみでなく、前日以前にユーザに出力した記事との間の類似度も求め、出力記事には、当日の記事だけからなる記事グループか、それ以前の日の記事も含まれているのかを区別するための情報を付加することが好ましい。これにより、利用者は、関連記事をさらに効率よく整理して読むことが可能となる。

【0029】

【発明の実施の形態】以下、図面を参照してこの発明の実施の形態について説明する。

【0030】まず、図1を参照して、この発明の情報フィルタリングシステム全体の構成について説明する。

【0031】情報フィルタリングシステムは、新聞社、通信社、または出版社などの複数の情報源2からテキストやイメージを含むテキスト記事の配信を受け、それを定期的に加入ユーザ端末3それぞれに送信する情報提供システムであり、このシステムの情報提供サービスは情報フィルタリングセンタ1によって実現されている。情報フィルタリングセンタ1は、通信網を介して複数の情報源2および複数の加入ユーザ端末3に接続された1つの計算機システムによって実現されており、ここには、情報フィルタリングのための制御や処理を行う中央処理装置4、プログラム並びにデータを格納する半導体メモリ、磁気ディスク、光ディスクなどの記憶装置5、回線や電波などの通信網介して情報源2からテキスト記事を受信する受信部6、回線や電波などの通信網介してユーザ端末3にテキスト記事を送信する送信部7などから構成されている。

【0032】各ユーザ端末3は、例えばパーソナルコンピュータやワークステーションなどの情報処理端末であり、情報フィルタリングセンタ1から送信されたテキスト記事を受信するテキスト情報受信部8と、受信したテキスト記事を画面表示する表示部9などを備えている。

【0033】情報フィルタリングセンタ1は、図2に示されているように、ユーザプロファイル10と称する一種の検索条件をユーザ毎に保持しており、そのユーザプロファイル10に従って該当するユーザに提供すべき記事を検索する。ユーザプロファイル10は、ユーザによって指定された複数のトピックなどから構成されてお

り、それらトピックに合致する記事が検索および選出されてユーザに送られる。次に、この情報フィルタリングセンタ1の具体的な構成について説明する。

【0034】(実施形態1)図3には、この発明の第1実施形態に係わる情報フィルタリングセンタ1の構成が示されている。図中、実線の矢印はデータの流れを示している。

【0035】情報フィルタリングセンタ1は、図示のように、ユーザプロフィール生成部11、ユーザプロフィール記憶部12、記事情報抽出部13、記事検索部14、記事選出部15、記事間類似度計算部16、提示情報生成部17、および記事情報記憶部18から構成されている。これら構成要素のうち、破線で囲まれているユーザプロフィール生成部11、記事情報抽出部13、記事検索部14、記事選出部15、記事間類似度計算部16、および提示情報生成部17は、例えば図1の中央処理装置14によって実行されるソフトウェアによって実現でき、またユーザプロフィール記憶部12および記事情報記憶部18は記憶装置5によって実現できる。

【0036】ユーザプロフィール生成部11は、予め各ユーザによって指定される要求・興味などを解析して検索のために必要なユーザプロフィールをユーザ毎に生成する。これらユーザプロフィールはユーザプロフィール記憶部12に記憶される。記事情報抽出部13は、各情報源2から到着したテキスト記事から、検索や記事同士の類似度計算に必要な情報を抽出し、それを生のテキスト記事と共に記事情報記憶部18に格納する。

【0037】記事検索部14は、各情報源2から到着した記事の中からユーザプロフィールに合致するものを検索する。この検索処理では、ユーザプロフィールと到着記事との間の類似度が調べられ、類似度の高い順に記事がソートされる。記事選出部15は、検索結果からユーザに提示する記事を選出するためのものであり、例えば類似度の値があるしきい値を越えた記事すべて、または類似度の高い上位のいくつかの記事が選択される。

【0038】記事間類似度計算部16は、記事間の類似度を調べるためのものであり、選択された記事同志の類似度を算出する。提示情報生成部17は、記事選出結果と記事間類似度計算結果をもとにユーザに提示する記事情報を生成する。記事情報記憶部18には、検索のための記事情報や記事間類似度計算結果などが記憶される。以下、これらユーザプロフィール生成部11、記事情報抽出部13、記事検索部14、記事選出部15、記事間類似度計算部16、および提示情報生成部17それぞれの具体的な処理内容について説明する。

【0039】図5には、ユーザプロフィール生成部11の処理の流れが示されている。

【0040】ユーザプロフィール生成部11は、入力として個々のユーザの要求・興味を受けつける(ステップS1)。ユーザの要求・興味は、「○○と××に関する

記事が読みたい」といった自然言語や、興味のあるトピック(話題)に頻出するキーワードの集合や、それらに優先順位や重みをつけたもの、または通常の文書検索における検索式のようなもので表される。

【0041】ユーザプロフィール生成部11は、これに対して単語辞書、同義語辞書などを利用して単語抽出、同義語展開などの言語処理を行い(ステップS2)、検索が可能となるような形式に変換してユーザプロフィールを作成する(ステップS3、S4)。作成されたユーザプロフィールはユーザ毎にユーザプロフィール記憶部12に記憶され、記事検索のための検索条件として利用される。

【0042】図5には、記事情報抽出部13の処理の流れの一例が示されている。

【0043】記事情報抽出部13は、入力として情報源から到着した記事を受けつけ(ステップS11)、これに対して文書解析用の辞書や情報抽出用の辞書を用いて、形態素解析、構文解析、書式解析などを行い、記事の情報源や日付、文字や単語その他の文書構成要素の頻度情報や出現位置、5W1H的な情報などの抽出を行う(ステップS12)。次いで、記事情報抽出部13は、これらの抽出された情報の集合体として記事を表現する(ステップS13)。例えば、出現した単語の頻度を要素とするベクトルにより記事を表現したり、5W1Hのテンプレートに実現値を代入したものにより表現したりする。このような記事の表現例をそれぞれ図6、図7に示す。図6は、その記事に出現した単語(半導体、メモリ、摩擦、不況、生産、…)の出現頻度(14、9、5、2、3)を要素とした頻度ベクトルであり、また図7は、情報源、文字数、記事見出し、トピック、日時、場所、主語、主動詞……などを項目とするテンプレートである。

【0044】記事情報抽出部13は、このようにして記事を表現した後、記事検索を高速に実現するための索引付け、つまりインデクシング処理も行い(ステップS14)、そしてベクトルやテンプレートで表現して記事およびインデクシング情報を記事情報記憶部18に記憶する(ステップS15)。

【0045】図8には、記事検索部14の処理の流れが示されている。

【0046】記事検索部14は、記事情報抽出部13によって抽出された記事情報を参照し、到着した記事の中からユーザプロフィールに適合するものを検索する。

【0047】これは、ユーザプロフィールと到着した記事の各々との類似度を算出することに相当する。この類似度は、検索の方式によって「ユーザプロフィールに適合する」「ユーザプロフィールに適合しない」などの離散値をとる場合もあるし、よく適合している記事ほど類似度の値が高くなるように連続値をとる場合もある。ここでは、より一般的である、類似度が連続値をとる場合

について説明する。

【0048】記事検索部14は、各ユーザのユーザプロフィールについて、以下の処理を行う。

【0049】まず、ユーザプロフィール記憶部12からプロフィールを読み込む(ステップS21)。次に、記事検索部14は、変数*i*に1を代入した後(ステップS22)、*i*番目の記事(1番目の記事)とユーザプロフィールとの類似度を計算する(ステップS23)。この類似度計算は、通常の検索処理に相当するもので、記事情報記憶部18に格納されている記事の表現や検索インデックスが参照される。

【0050】次いで、記事検索部14は、変数*i*の値を+1更新した後、そのときの*i*の値が到着記事数よりも大きいかなかを調べ(ステップS24、S25)、大きくない場合には類似度計算されてない記事が残っていると認識し、*i*の値が到着記事数よりも大きくなるまで、ステップS23～S25を繰り返す。到着した全ての記事に対してユーザプロフィールとの類似度の計算が終わると、すなわち、到着した全ての記事を検索対象とした検索処理が終わると、記事検索部14は、到着した記事をユーザプロフィールとの類似度が高い順にソートし、記事のランキングを行う(ステップS26)。このランキングの結果は、記事情報記憶部18に記憶される。ランキング結果の一例を図9に示す。

【0051】図10に、記事選出部15の処理の流れを示す。

【0052】記事選出部15は、記事検索部14により検索されランキングされた到着記事を記事情報記憶部18から読み込み(ステップS31)、その中から、実際にユーザに提示するものを選出する(ステップS32)。ユーザに提示することが決まった記事の情報は、再び記事情報記憶部18に格納される。

【0053】記事の選出の方法としては、例えば、ユーザに提示する記事の件数*N*をユーザ側があるいはセンタ側が予め決めておき、ランキング上位*N*件を提示することにしたり、あるいは、ユーザプロフィールとの類似度がある閾値以上の記事を提示することにするなどが考えられる。図11は、図9のようなランキング結果が得られている場合にその上位10件を選出した例を示している。

【0054】また図12は、図9のようなランキング結果が得られている場合にユーザプロフィールとの類似度が0.86以上の記事を選出した例を示している。

【0055】さらに、図13は、一人のユーザに対して複数の検索及びランキングが行われた場合において、これら複数のランキング結果の上位部分をマージしてユーザに提示する記事を選出する例を示している。

【0056】この例では、「半導体技術」、「低価格パソコン」、「人工知能」の3つのトピックに関する検索が別個に行われており、3つのランキング結果の上位か

ら、記事A1、B1、C1、A2、B2が選出されている。

【0057】記事A1及びA2はトピック「半導体技術」に適合したもの、記事B1及びB2はトピック「低価格パソコン」に適合したもの、そして記事C1はトピック「人工知能」に適合したものである。

【0058】ここでの記事の選出の方法としても、図11のように一定件数を選出する、図12のように類似度が一定値以上の記事を選出するなどが考えられる。

10 【0059】図14には、記事間類似度計算部16の処理の流れが示されている。

【0060】記事検索部14が、ユーザプロフィールと記事との類似度を計算する、換言すれば、ユーザプロフィールを検索式とし、記事を検索対象とした通常の実行を行うのに対し、記事間類似度計算部16は、記事同士

の類似度を計算する。

【0061】類似度計算は、例えば図6や図7のような記事の表現同士を比較することにより行われ、計算結果は記事情報記憶部18に記憶される。

20 【0062】ここでは、新聞社などの記事の情報源2は複数存在するものとし、記事間類似度計算の対象となるのは、異なる情報源から到着した記事同士、例えば、新聞社Mから到着した記事と新聞社Nから到着した記事とする。

【0063】異なる情報源から到着した記事同士の全ての組合せについて記事間類似度を計算してもよいが、ここでは、記事選出部15により選出された記事に対してのみ記事間類似度を計算するという計算コストの低い方法について説明する。

30 【0064】すなわち、記事間類似度計算部16は、まず、記事選出部15により選出された記事を記事情報記憶部18から読み込む(ステップS41)。次いで、記事間類似度計算部16は、読み込んだ記事のうち、異なる情報源から到着したものの同志の類似度を計算し、その結果を記事情報記憶部18に格納する(ステップS42)。

【0065】以下、記事間類似度計算の具体例を説明する。

40 【0066】図15には、記事選出部15により選出された、異なる情報源から到着した記事の例が示されている。この例では、記事A～Dの4つの記事がユーザに提示されることになっている。

【0067】記事A及びDは新聞社Mから到着した記事、記事Bは新聞社Nから到着した記事、そして記事Cは出版社Oから到着した記事である。

50 【0068】この場合、記事Aと記事B、記事Aと記事C、記事Bと記事C、記事Cと記事Dの組合せについて記事間類似度が計算される。記事Aと記事Dは同一の情報源から到着した記事であるため、類似度計算を行わない。

【0069】図16には、提示情報生成部17の処理の流れが示されている。

【0070】提示情報生成部17は、記事情報記憶部18から、記事選出部15により選出された記事の情報及び記事間類似度計算部16により計算された記事間類似度を読み込む（ステップS51、S52）。

【0071】そして、提示情報生成部17は、互いに類似度が高く、かつ情報源が異なる記事の集合を、重複記事の集合として分類する（ステップS53）。ここで、重複記事とは、同じ出来事に対して複数の情報源が独自に作成した記事のことであり、内容的には同一あるいはほぼ同じと考えてよい記事をいう。

【0072】この後、提示情報生成部17は、重複記事の提示を回避するために、重複記事集合から、代表としてユーザに提示する記事をひとつ、一般にはN個選択する（ステップS54）。そして、提示情報生成部17は、選択した記事の本文に対して、選択しなかった記事の情報を関連記事情報として付加することなどにより、ユーザに提示する情報を生成し、それを出力する（ステップS55、S56）。

【0073】ここで、重複記事と関連記事情報の具体例を説明する。

【0074】図17は、一件のプレスリリースから重複記事が派生する例を示している。ある出来事について情報を記したプレスリリース記事Pが、新聞社M、N、Oに送られると、各新聞社はこれを編集し、コメントを加えるなどして、独自の記事M、N、Oを作成する。記事M、N、O、Pが各情報源から情報フィルタリングセンターに送られるとすると、記事M、N、O、Pは重複記事となる。

【0075】また、図18は、一件の出来事から重複記事が作成される例を示している。

【0076】この例では、新聞社M、N、Oが同一の出来事に対して独自に取材を行い、記事M、N、Oが作成されている。これらが情報フィルタリングセンタ1に送られるとすると、記事M、N、Oは重複記事となる。

【0077】情報フィルタリングの本来の主旨は、膨大な情報の中にある所望の情報をできるだけ効率よくユーザにアクセスさせることであるから、一般には、ユーザに提示する記事に重複記事が多く含まれていることは好ましくないと考えられる。例えば、図18の例において、ユーザに記事M、N、Oを全て提示してしまうと、ユーザはひとつの出来事についての情報を得るのに3つの記事を読まなければならないになってしまう。

【0078】提示情報生成部17は、以上のような重複記事の提示を回避するために、重複記事集合から、代表としてユーザに提示する記事をひとつ、一般にはN個選択する。以後、ひとつだけ選択する場合についてのみ説明する。

【0079】図19に、図15の4つの記事に対して記

事間類似度計算を行った結果得られる重複記事集合の例を示す。

【0080】この例では、記事AとC、記事BとDの類似度が高かったため、ふたつの重複記事集合が得られている。

【0081】提示情報生成部17は、一定の戦略に従って、各重複記事集合から記事をひとつずつ選択する。

【0082】例えばユーザ側あるいはサービスセンタ側が新聞社Mを最優先するように予め決めておいたとすると、最終的にユーザに提示される記事は新聞社Mから到着した記事A、Dとなる。

【0083】同様に、一般には情報量の一番多いプレスリリースを最優先して選択することも考えられる。

【0084】また、検索結果のランキングで最も上位にあったものを選択することも考えられる。

【0085】例えば図19において、ユーザプロファイルと記事との類似度は重複記事集合1では記事Cが高く、重複記事集合2では記事Dが高いので、最終的にユーザに提示される記事は記事C、Dとなる。

【0086】さらに、記事の長さが最も長い、あるいは最も少ないものを選択するなどの戦略も考えられる。

【0087】これまで述べてきた処理により、ユーザに提示する記事の候補の中から重複記事が排除される。最終的に排除された重複記事に関する情報は、各記事の本文情報に付加してユーザに提示される。

【0088】図20に、排除された重複記事に関する情報を記事の本文情報に付加して提示する例を示す。

【0089】この例では、ユーザに提示されている記事の本文情報に加えて、この記事と内容が同一と判断された他の情報源の記事に関する情報が付加情報として与えられている。具体的には、記事の見出しや情報源、文字数、そして現在本文が提示されている記事との類似度などがリストアップされている。

【0090】この例では、「〇×社が情報サービス事業から撤退した」という内容の記事が〇〇新聞社、△△新聞社、□□新聞社の3つの情報源から得られたが、ユーザに提示する記事としては〇〇新聞社の記事が選ばれたということになる。

【0091】このように排除された重複記事に関する情報を記事の本文情報に付加して提示すると、内容的には同じだが情報源が異なる記事を何度も読むことを回避でき、かつ各情報源が同一の出来事に対してどのように報道しているかの概観を得ることができると考えられる。

【0092】図21には、図20の関連情報提示形態の変形例が示されている。

【0093】すなわち、図20はべたテキストとして関連情報をユーザ端末に表示したが、図21では、付加情報のテキスト部分がハイパーテキストなどで構造化されており、これを利用して排除された重複記事の本文にアクセスすることを可能にしている。

【0094】この例では、付加情報のエリアにおける記事見出しがマウスなどの装置により選択できるボタンになっており、ユーザは関連記事1を選択することによって、その関連記事1の本文を参照することができる。

【0095】図22及び図22に、図21において関連記事1が選択された場合に関連記事1の本文を表示する例を示す。

【0096】図21において本文が表示されていた「半導体協議の…」という記事は、図22においては付加情報のエリアに見出しなどの情報のみが表示されており、その代わりに本文情報のエリアには関連記事1の本文が表示されている。

【0097】図22から図21の状態に戻すためには、ユーザは図22の付加情報のエリアにおける「半導体協議の…(元の記事)」というボタンを選択すればよい。

【0098】また図23においては、図21で表示されている情報を保持しながら、関連記事1の本文情報を新たに開いたウィンドウ上に表示している。このような表示方法をとれば、複数の重複記事を比較することが可能となる。

【0099】図21の画面から図22の画面への移行は、図24の処理の流れに従って次のように実行される。

【0100】提示情報生成部17は、図21に示されているように提示記事の本文情報に関連記事の情報を付加してユーザ端末3に画面表示させる(ステップS61)。次いで、関連記事のボタンが選択されるというイベントが発生すると、提示情報生成部17は、選択された関連記事の本文情報を記事情報記憶部18から取り出し(ステップS62、S63)、図22に示されているように、元の記事の情報を付加情報のエリアに、選択された関連記事の本文を本文情報エリアに表示する(ステップS64)。

【0101】なお、このような画面切り替えは、センタ1からユーザ端末3に予め関連記事の本文情報を送信しておけば、ユーザ端末3側の制御で行うこともできる。

【0102】図21の画面から図23の画面への移行は、図25の処理の流れに従って次のように実行される。

【0103】提示情報生成部17は、図21に示されているように提示記事の本文情報に関連記事の情報を付加してユーザ端末3に画面表示させる(ステップS71)。次いで、関連記事のボタンが選択されるというイベントが発生すると、提示情報生成部17は、選択された関連記事の本文情報を記事情報記憶部18から取り出し(ステップS72、S73)、図23に示されているように、選択された関連記事の本文をウィンドウ表示する(ステップS74)。

【0104】この画面切り替えについても、センタ1からユーザ端末3に予め関連記事の本文情報を送信してお

けば、ユーザ端末3側の制御で行うことができる。

【0105】また、図20や図21のように付加情報のエリアに付加する関連記事は、前述の重複記事集合から記事を選択するのと同様な戦略により絞り込みを行ってもよい。

【0106】図21～23のように、重複記事集合を代表して本文が表示されている記事から、それ以外の重複記事の本文にアクセスできるようにすると、情報フィルタリングシステムによって選択された代表記事が不適切なものであった場合にも、ユーザが他の重複記事を選択して読むことが可能となる。

【0107】例えば、情報フィルタリングシステムがユーザの希望によりN新聞の記事を優先して選択する戦略をとっていたとしても、ユーザがある出来事に関してだけはN新聞の記事ではなくプレスリリースを読みたいといった場合に有効である。

【0108】また、同一の出来事に対する複数の新聞社の見解などを比較することも可能となる。

【0109】図26には、記事の重複が起こっている場合に、ユーザに提示する記事の一覧表を重複記事情報とともに表示した例が示されている。

【0110】この例では、ユーザに提示する記事は4件あるが、そのうちの3件目である「〇×社 情報サービスビジネスから撤退」という記事には重複記事が2件存在する。

【0111】各記事の見出しの後にはユーザプロフィールと記事との類似度の値が表示されているが、重複記事に関してはそれとは別に、元の記事と重複記事との類似度の値も表示されている。これは重複記事である確からしさを示していると言える。ここで、元の記事とは「〇×社 情報サービスビジネスから撤退」の記事などを指す。

【0112】以上の説明では、ひとつのユーザプロフィールに対する処理を中心に述べてきた。

【0113】一般には、情報フィルタリングサービスを受けるユーザは複数いるので、情報フィルタリングセンタは、ユーザ毎にユーザプロフィールを保持し、フィルタリングの各処理を行うことになる。

【0114】(実施形態1の変形例1)次に、記事間類似度計算部16及び提示情報生成部17の他の構成例について説明する。

【0115】図27に、記事間類似度計算部16の処理の流れを示す。

【0116】記事検索部14が、ユーザプロフィールと記事との類似度を計算する、換言すれば、ユーザプロフィールを検索式とし、記事を検索対象とした通常の検索を行うのに対し、記事間類似度計算部16は、記事同士の類似度を計算する。

【0117】類似度計算は、例えば図6や図7のような記事の表現同士を比較することにより行われ、計算結果

は記事情報記憶部18に記憶される。

【0118】ここでは、N回前までの情報フィルタリングで得られた記事情報が記事情報記憶部18に保存されているものとする。

【0119】例えば、情報フィルタリングサービスが一日一回行われ、かつNを1とした場合、昨日の情報フィルタリングで得られた記事情報は保存されているということを意味する。以後、主にN=1として説明する。

【0120】このシステムでは、記事間類似度計算の対象となるのは、今回到着した記事と、前回までに到着した記事との組である。

【0121】今回到着した記事と、前回までに到着した記事との全ての組合せについて類似度計算を行ってもよいが、以後は、より計算コストの低い方式、つまり記事選出部により今回選出された記事と、前回までにユーザに提示された記事との組合せについての類似度計算を行う場合について説明する。

【0122】すなわち、記事間類似度計算部16は、まず、記事選出部15により選出された記事の情報を記事情報記憶部18から読み込み、次いで、前回までのフィルタリングでユーザに提示した記事の情報を記事情報記憶部18から読み取る(ステップS81、S82)。そして、記事間類似度計算部16は、記事選出部15により今回選出された記事と、前回までにユーザに提示された記事との組合せについて類似度計算を行い、その結果を記事情報記憶部18に格納する(ステップS83)。

【0123】図28には、今回記事選出部15により選出された記事の集合と、前回ユーザに提示された記事の集合の例を示す。

【0124】この例では、前回はユーザには記事A、B、C、Dが提示されており、今回は記事E、F、G、Hが提示されようとしている。

【0125】この場合、類似度計算は、記事Aと記事E、記事Aと記事Fといったように $4 \times 4 = 16$ の組合せについて計算されることになる。

【0126】また、この変形例として、一定の条件を満たす記事のみ類似度計算の対象としてもよい。

【0127】例えば、図28において、情報源が同じ記事同士の類似度のみを計算することになると、今回新聞社Mから到着した記事Eに関する類似度計算は、前回新聞社Mから到着した記事A、Bのみについて行えばよいことになる。

【0128】また、例えば、図28において、ユーザプロフィールとの類似度が一定値以上の記事のみを類似度計算の対象とすることも考えられる。

【0129】ユーザプロフィールとの類似度が0.8以上の記事のみを対象とすると、記事Eと記事A、記事Cと記事Aの組合せのみ計算すればよいことになる。

【0130】図29には、提示情報生成部17の処理の流れが示されている。

【0131】提示情報生成部17は、記事情報記憶部18から、今回記事情報選出部15により選出された記事の情報及び前回までにユーザに提示された記事の情報及び記事間類似度計算部16により計算された記事間類似度を読み込む(ステップS91～S93)。そして、今回の記事の本文情報を前回までの関連記事の情報とともにユーザに提示する(ステップS94、S95)。

【0132】図30及び図31には、今回の記事の本文情報に前回までの関連記事の情報を付加して提示する例が示されている。

【0133】図30では、ユーザに今回初めて提示された「半導体協議の…」という記事の本文情報に加えて、半導体に関する昨日までの記事の情報が付加情報として与えられている。具体的には、前回までの記事の見出しや情報源、文字数、そして今回提示された記事との類似度などがリストアップされている。

【0134】この例では、今回提示されたのは〇〇新聞の15日付の記事であり、前回までの関連記事としては〇△新聞及び〇〇新聞の14日付の記事が表示されている。

【0135】また図31では、ユーザに今回提示された「シリーズ：半導体摩擦(その3)」という記事の本文情報に加えて、同じ〇〇新聞社から昨日までに到着した「シリーズ：半導体摩擦(その1)」及び「シリーズ：半導体摩擦(その2)」という記事に関する情報が表示されている。

【0136】また、実施形態1で示した図21～23は、この例における図30及び図31の変形例にもなっている。

【0137】すなわち、このシステムでも、実施形態1と同様に、ユーザが前回までの関連記事の本文にアクセスできるようにすることが考えられる。

【0138】図21～23では、本文情報と付加情報が完全に分離されているが、本文情報中に前回までの記事情報を埋め込んで提示することも考えられる。

【0139】図32に、今回の記事の本文情報中に前回までの関連記事の情報を埋め込んで提示する例を示す。

【0140】この例では、「〇〇沖の××地震 再び活発化」という19日付けの記事の本文が表示されているが、その第一文の「〇〇県〇〇沖で先月14日から始まった××地震は…」の一部がマウスなどで選択できるボタンになっている。

【0141】ユーザがこれを選択すると、前回までの記事のうちこの文と類似した情報を含む記事に関する情報が表示される。

【0142】図33は、図31においてユーザが第一文を選択した場合に、その文と関係の深い前回までの記事のリストを表示した例である。

【0143】この例では、「〇〇沖で地震 マグニチュード4」などの、14日付の記事の見出しや情報源、文

字数、今回の記事との類似度などがリストアップされている。

【0144】図34は、図33においてユーザが「〇〇沖で地震 マグニチュード4」という関連記事を選択した場合に、その記事の本文を表示した例である。

【0145】また、図32においてユーザが第一文を選択した直後に、図34のように関連記事の本文をひとつ以上表示してもよい。

【0146】図32のように、今回の記事の本文情報中に前回までの関連記事の情報を埋め込んで提示するという形態を実施するためには、今回の記事と前回までの記事との類似度を計算するかわりに、今回の記事の本文の各構成要素と前回までの記事との類似度を計算することになる。

【0147】本文の構成要素としては、段落、文、節、句、単語、などが考えられる。

【0148】また、これをさらに変形して、前回までの関連記事情報も記事単位ではなく本文の構成要素単位で提示してもよい。

【0149】例えば、図34のように関連記事の全文を表示するかわりに、第一段落のみを表示することが考えられる。

【0150】以上のように、今回提示された記事からそれと関係のある前回までの記事にアクセスできるようにすれば、時間の経過とともに状況が変わっていくような出来事についてその経緯を把握することや、連載記事など複数の記事にわたる情報を把握することが容易になる。

【0151】さらに、今回の記事を読んだときに、過去に提示された記事を思い出し、その内容を再確認したくなった場合などにも有効である。

【0152】（実施形態1の変形例2）次に、記事間類似度計算部16及び提示情報生成部17のさらに他の構成例について説明する。

【0153】図35には、記事間類似度計算部16の処理の流れが示されている。

【0154】記事検索部14が、ユーザプロフィールと記事との類似度を計算する、換言すれば、ユーザプロフィールを検索式とし、記事を検索対象とした通常の検索を行うのに対し、記事間類似度計算部16は、記事同士の類似度を計算する。

【0155】類似度計算は、例えば図6や図7のような記事の表現同士を比較することにより行われ、計算結果は記事情報記憶部18に記憶される。

【0156】この例において記事間類似度計算の対象となるのは、今回到着した記事同士の組合せである。

【0157】到着した記事全てについて類似度計算を行ってもよいが、以後、より計算コストの低い、今回記事選出部15により選出された記事同士についてのみ類似度計算を行う場合について説明する。

【0158】今回の記事同士について類似度計算を行う点では、実施形態1と同じであるが、実施形態1が情報源の異なる記事間に対して計算を行っていたのに対し、ここではそのような限定はしていない。

【0159】図15のように記事選出部15により4つの記事が選出された場合、記事間類似度計算部16は、それら記事を記事情報記憶部18から読取り（ステップS101）、記事Aと記事B、記事Aと記事C、記事Aと記事D、記事Bと記事Dといったように全ての組合せについて類似度計算を行う（ステップS102）。

【0160】また、一定の条件を満たす記事のみ類似度計算の対象としてもよい。

【0161】図36には、提示情報生成部17の処理の流れが示されている。

【0162】提示情報生成部17は、記事情報記憶部18から、記事情報選出部15により選出された記事の情報及び記事間類似度計算部16により計算された記事間類似度を読み込む（ステップS111、S112）。そして、提示情報生成部17は、今回の記事の本文情報を今回の他の関連記事の情報とともにユーザに提示する（ステップS113、S114）。

【0163】図37に、今回の記事の本文情報を今回の他の関連記事の情報とともに提示する例を示す。

【0164】この例では、15日付の「半導体協議の…」という記事の本文情報に加えて、同じ15日付けの半導体に関する記事の情報が付加情報として与えられている。これにより、実施形態1における重複記事が表示されてしまう可能性があるが、このような場合には実施形態1の重複記事削除処理を行ってもよい。

【0165】また、図37の付加情報のエリアに表示されている「××社 半導体シェア独占…」という記事の本文情報を見るときには、図38のように「半導体協議の…」という記事が付加情報のエリアに表示されることになる。

【0166】また、実施形態1で示した図21～23は、この例の図37及び図38の変形例にもなっている。

【0167】すなわち、実施形態1と同様に、ユーザが当日の関連記事の本文に直接アクセスできるようにすることが考えられる。

【0168】（記事間類似度の記事提示順序への反映）これまでの、主に個々の記事をユーザに提示する際の関連記事情報の付加について述べてきたが、今回の記事同士の記事間類似度を利用して、ユーザに提示する記事の順序を決定することも可能である。

【0169】図39に、記事間類似度を記事の提示順序に反映させる例を示す。

【0170】この例では、ユーザプロフィールは半導体技術、低価格パソコン、人工知能という3つの異なる分野に関する語の集合であるとする。

【0171】これにより検索を行うと、図39(a)のように、3つの異なる分野の記事が混在した検索結果が得られる。

【0172】ここで、例えば上位8件、あるいはユーザプロフィールとの類似度が0.80以上の記事を選択し、そのままの順序でユーザに提示すると、ユーザは半導体、低価格パソコン、人工知能、半導体、低価格パソコン、のような順序で記事を読むことになってしまう場合がある。

【0173】ユーザプロフィールとの類似度が近い順に記事を読むことが有効な場合もあると考えられるが、このように複数分野の記事が混在している場合には、図39(b)のように、内容の類似した記事を集め、それをひとかたまりにして表示した方がユーザにとって分かりやすいと考えられる。

【0174】この例では、半導体に関する記事が初めの3件、低価格パソコンに関する記事が次の3件、そして残りの2件が人工知能に関するものとなっている。

【0175】以上のように、この実施形態1のシステムでは、頻度ベクトルなどを使用して記事間でその表現を比較することによって記事同志の類似度が算出され、その類似度に従って、ユーザに提示される記事に関連する関連記事が決定される。この関連記事の情報は、ユーザに提示される記事の本文情報に付加されてユーザに送られる。類似度計算を行う対象としては、今回提示されている記事同士、または今回到着した記事と前回までに到着した記事との間で行うことが好ましい。これにより、今回提示されている記事同士の関係や、今回提示されている記事と過去のフィルタリングで提示された記事との関係が明確になり、記事同士の関連性などをユーザに知らせることができる。

【0176】また、記事同志の類似度を算出することによって重複記事の存在を調べれば、重複記事の本文情報についてはユーザに提示せずに、その重複記事の見出しなどの情報だけを関連記事情報として付加してユーザに提示することもできる。これにより、例えば異なる複数の情報源から得られた同一内容に関する記事が、重複してユーザに提示されることを自動的に回避することができる。

【0177】よって、一回の情報フィルタリングでユーザに複数の記事を提示する際、記事間の関係を明確にして提示できるようになり、ユーザにとって記事内容の理解が容易になると考えられる。

【0178】(実施形態2)次に、この発明の情報フィルタリングシステムの第2の実施形態について説明する。このシステム全体の構成は、図1と同様であり、ユーザ毎にユーザプロフィールが保持しており、そのユーザプロフィールを利用して記事の検索が行われる。ここで、ユーザプロフィールとは、前述したように、ユーザの関心が高いトピックに適合する記事を検索するための

検索条件のことをいう。

【0179】図40に、実施形態2で使用されるユーザプロフィールの概念図を示す。

【0180】この例では、あるユーザAは「半導体技術」、「半導体貿易」という2つのトピックを選択している。また別のユーザBは、「半導体貿易」、「低価格パソコン」、「人工知能」という3つのトピックを選択している。

【0181】このとき、ユーザAのユーザプロフィールは、「半導体技術」に関する記事を検索するための検索条件及び「半導体貿易」に関する記事を検索するための検索条件から構成される。同様に、ユーザBのユーザプロフィールは、「半導体貿易」に関する記事の検索条件、「低価格パソコン」に関する記事の検索条件、及び「人工知能」に関する記事の検索条件から構成される。

【0182】図41には、実施形態2に係わる情報フィルタリングセンタ1の構成が示されている。情報フィルタリングセンタ1は、図示のように、ユーザプロフィール生成部21、トピック記憶部22、記事情報抽出部23、記事検索部24、記事選出部25、付加情報生成部26、および記事情報記憶部27から構成されている。これら構成要素のうち、破線で囲まれているユーザプロフィール生成部21、記事情報抽出部23、記事検索部24、記事選出部25、付加情報生成部26は、例えば図1の中央処理装置14によって実行されるソフトウェアによって実現でき、またトピック記憶部22および記事情報記憶部27は記憶装置5によって実現できる。

【0183】ユーザプロフィール生成部21は、入力として個々のユーザの要求・興味を受けつける。ユーザの要求・興味は、「〇〇と××に関する記事が読みたい」といった自然言語や、興味のあるトピックに頻出するキーワードの集合や、それらに優先順位や重みをつけたもの、または通常の文書検索における検索式のようなもので表される。

【0184】ユーザプロフィール生成部21は、これに対して単語抽出、同義語展開などの言語処理を行い、検索が可能となるような形式に変換してユーザプロフィールを作成する。ユーザプロフィールはユーザ毎にトピック記憶部22に記憶される。また、ユーザプロフィール生成部21は、既にユーザに送信された各記事がユーザにとって有用であったか否かなどの情報についてユーザからフィードバックを受け、その情報を反映してトピック記憶部22の検索条件を修正するというレlevanceフィードバック機能も有している。

【0185】記事情報抽出部23は、入力として情報源から到着した記事を受けつけ、これに対して形態素解析、構文解析、書式解析などを行い、記事の情報源や日付、文字や単語その他の文書構成要素の頻度情報や出現位置、5W1H的な情報などの抽出を行う。そして、記事をこれらの抽出された情報の集合体として表現する。

例えば、出現した単語の頻度を要素とするベクトルにより記事を表現したり、5W1Hのテンプレートに実現値を代入したものにより表現したりする。このような記事の表現例は、それぞれ図6および図7で説明した実施形態1のものと同じである。

【0186】記事情報抽出部23は、記事検索を高速に実現するためのインデクシング処理も行う。記事情報抽出部23により抽出された記事情報は、記事情報記憶部27に記憶される。

【0187】次に、図42を参照して、記事検索部24の処理の流れを説明する。

【0188】記事検索部24は、トピック記憶部22に記憶されている各トピックの検索条件と、記事情報抽出部23によって抽出された記事情報とを参照し、各トピックに適合する到着記事を検索する。これは、トピックと到着記事との類似度を算出することに相当する。この類似度は、検索の方式によって「トピックに適合する」「トピックに適合しない」などの離散値をとる場合もあるし、よく適合している記事ほど類似度の値が高くなるように連続値をとる場合もあるが、ここでは、より一般的である類似度が連続値をとる場合について説明する。

【0189】記事検索部24は、各トピックについて、以下の処理を行う。

【0190】まず、記事検索部24は、変数*i*に1を代入した後（ステップS121）、*i*番目のトピック（トピック1）の検索条件をトピック記憶部22から取り出す（ステップS122）。この後、記事検索部24は、変数*j*に1を代入した後（ステップS123）、トピック1（トピック1）と到着記事*j*（到着記事1）との類似度を計算し、満足された検索条件の情報と共に類似度を記事情報記憶部27に格納する（ステップS124）。この類似度計算は、通常の検索処理に相当するもので、記事情報記憶部18に格納されている記事の表現や検索インデックスが参照される。

【0191】次いで、記事検索部24は、変数*j*の値を+1更新した後、そのときの*j*の値が到着記事数よりも大きいか否かを調べ（ステップS125、S126）、大きくない場合には類似度計算されていない記事が残っていると認識し、*j*の値が到着記事数よりも大きくなるまで、ステップS124～S126を繰り返す。到着した全ての記事に対してトピック1との類似度の計算が終わると、記事検索部24は、到着した記事をユーザプロフィールとの類似度が高い順にソートし、記事のランキングを行う（ステップS127）。このランキングの結果は、記事情報記憶部27に記憶される。

【0192】この後、記事検索部24は、変数*i*の値を+1更新した後、そのときの*i*の値が全トピック数よりも大きいか否かを調べ（ステップS128、S129）、大きくない場合には類似度計算されていないトピックが残っていると認識し、*i*の値が全トピック数よりも

大きくなるまで、ステップS122～S129を繰り返す。

【0193】図43は、記事検索部24によりランキングされたトピック*i*に対する到着記事の概念図を示す。このように、到着記事はトピック毎にランキングされる。

【0194】図44には、記事選出部25の処理の流れが示されている。

【0195】記事選出部25は、記事検索部24により記事情報記憶部27に格納された各トピックの検索結果の中から、各ユーザに提示する記事を選出する。

【0196】すなわち、まず、記事選出部25は、変数*i*に1を代入した後（ステップS131）、ユーザ*i*（ユーザ1）のユーザプロフィールをトピック記憶部22から取り出す（ステップS132）。この後、記事選出部25は、変数*j*に1を代入した後（ステップS133）、ユーザ*i*のトピック*j*（トピック1）の検索結果の記事情報記憶部27から取り出し、その中からユーザに提示する記事を選出する（ステップS135）。記事の選出の方法としては、例えば、ユーザに提示する記事の件数*N*をユーザ側があるいはセンタ側が予め決めておき、ランキング上位*N*件を提示することにしたり、あるいは、ユーザプロフィールとの類似度がある閾値以上の記事を提示することにするなどが考えられる。選出された記事の情報は、記事情報記憶部27に格納される。

【0197】次いで、記事選出部25は、変数*j*の値を+1更新した後、そのときの*j*の値がユーザ*i*の指定したトピック数よりも大きいか否かを調べ（ステップS136、S137）、大きくない場合には選出されていない他のトピックの検索結果が残っていると認識し、*j*の値がユーザ*i*のトピック数よりも大きくなるまで、ステップS134～S137を繰り返す。ユーザ*i*のすべてのトピックに対しての記事選出が終わると、記事選出部25は、変数*i*の値を+1更新した後、そのときの*i*の値が全ユーザ数よりも大きいか否かを調べ（ステップS138、S139）、大きくない場合には記事選出されていないユーザが残っていると認識し、*i*の値が全ユーザ数よりも大きくなるまで、ステップS132～S139を繰り返す。

【0198】このような処理により、例えば、図45のように、「半導体貿易」、「低価格パソコン」、「人工知能」という3つのトピックを選択しているユーザに対しては、「半導体貿易」の検索結果、「低価格パソコン」の検索結果、「人工知能」の検索結果の3つが取り出され、これらのうち上位の記事の中からユーザに提示するものが選出される。

【0199】図46には、付加情報生成部26の処理の流れが示されている。

【0200】付加情報生成部26は、全てのユーザに対して以下を行う。

【0201】まず、付加情報生成部26は、変数*i*に1を代入した後(ステップS141)、ユーザ*i*(ユーザ1)のユーザプロフィールをトピック記憶部22から取り出す(ステップS142)。次に、付加情報生成部26は、記事選出部25によって選出されたユーザ1に提示する記事と、これらの記事が満足した検索条件に関する情報とを記事情報記憶部27から取り出す(ステップS143)。

【0202】ここで、記事が満足した検索条件に関する情報とは、その記事がユーザの選択したトピックのうちいずれに適合したか、トピックの検索条件の中のどのような条件に適合したか、などの情報をいう。検索条件とは、どのような言語表現が記事中のどのような位置に、どのような頻度で含まれていたか、記事の主題・行為やその動作主は何かなどの記事が満たすべき条件を、通常の文書検索で用いるプール式や自然言語、その他の記事検索部により処理が可能な形式で記述したものをいう。

【0203】この後、付加情報生成部26は、記事選出部25によって選出された記事に、これらの記事が満足した検索条件に関する情報を付加してユーザ*i*に提示する(ステップS144)。そして、付加情報生成部26は、変数*i*の値を+1更新した後、そのときの*i*の値が全ユーザ数よりも大きいか否かを調べ(ステップS145、S146)、大きくない場合には付加情報が生成されていないユーザが残っていると認識し、*i*の値が全ユーザ数よりも大きくなるまで、ステップS142~S146を繰り返す。

【0204】図47には、あるユーザのために選出された記事の記事見出しのリストに、各記事が適合したトピックの情報を付加してそのユーザに提示した表示例が示されている。

【0205】ここでは、ユーザは「半導体貿易」、「低価格パソコン」、「人工知能」という3つのトピックを選択しているものとする。

【0206】この例では、ユーザに6つの記事の記事見出しが提示されており、これらの記事のうち3つが「半導体貿易」に適合した記事、2つが「低価格パソコン」に適合した記事、そして残りの1つが「半導体貿易」と「低価格パソコン」の両方に適合した記事になっている。

【0207】このように、ひとつの記事が複数のトピックに適合する場合があっても、その記事が提示された根拠が表示される。

【0208】また、この例では、各行の最後の欄に、記事検索部24により検索時に計算された、適合したトピックと記事との類似度の値が表示されている。

【0209】記事番号6の記事は、2つのトピックに適合したため、「半導体貿易」との類似度は1.05、「低価格パソコン」との類似度は0.80、というように2つの類似度が表示されている。

【0210】図48に、図47と同じユーザに対して各トピックに適合した記事の件数情報を提示した表示例を示す。

【0211】図48(a)では、ユーザの選択している各トピックに適合した記事の件数情報を表形式で表示している。

【0212】「半導体貿易」に適合した記事は、図47における記事番号1、2、3、及び6の記事であるので、記事数は4と表示されている。同様に、「低価格パソコン」に適合した記事は、図47における記事番号4、5、及び6の記事であるので、記事数は3と表示されている。また、この例では「人工知能」に適合する記事はないので、記事数は0となっている。

【0213】ユーザに提示された記事数は、「半導体貿易」の4件と「低価格パソコン」の3件のうち1件重複があるので6件となっている。

【0214】また、この変形例として、図47の記事番号6のように複数のトピックに適合した記事の件数は別個にカウントするようにしてもよい。

【0215】この場合、例えば図48(a)の「半導体貿易」の件数は、このトピックのみに適合した記事の件数という意味で3件となる。

【0216】図48(b)では、ユーザの選択している各トピックに適合した記事の件数情報をベン図形式で表示している。

【0217】この例では、図47における記事番号1、2、3の3つの記事が「半導体貿易」のみに適合した記事であり、記事番号4、5の2つの記事が「低価格パソコン」のみに適合した記事であり、記事番号6の記事が両方に適合した記事であることが明示されている。

【0218】この例では、図48(a)に比べ、各トピックの適合件数と全記事数との関係がより明確になっている。

【0219】図49に、あるユーザのために選出された記事の要約文・抜粋文あるいは本文をトピック別にまとめてそのユーザに提示した表示例を示す。

【0220】ここで、要約文とは、もとの記事の本文を加工して要点がつかめるようにしたテキストをいい、抜粋文とは、もとの記事の本文の一部を加工せずに抜き出したテキストをいう。

【0221】この例では、「半導体貿易」に関する3つの記事が並べて最初に表示されており、その後に「低価格パソコン」に関する記事が続いている。

【0222】以上の説明のように、ユーザに提示する各記事がどのトピックに適合したのかを明示することにより、ユーザは記事の内容理解や、どの記事を読み、どの記事を読まないかなどの判断が容易になり、より効率的な情報収集ができると考えられる。

【0223】図50に、記事が満足した検索条件に関する情報を記事本文のヘッダ情報として付加してユーザに

提示した表示例を示す。

【0224】この例では、表示中の記事が、ユーザの選択しているトピックのうち「半導体貿易」に適合したものであることが「該当トピック」の行に明示されている。

【0225】その下には、「半導体貿易」と記事との類似度が1.32であつたことが表示されている。

【0226】さらに、「半導体貿易」に関する記事を検索するために用いられた検索条件と、これらの条件のうち表示中の記事が満足したものが並べて表示されている。

【0227】また、図50の本文中では、テキストの一部が強調表示されている。

【0228】ここで、強調表示とは、アンダーラインなどの付加記号を伴った表示、異なる字体や大きさの文字による表示、異なる色による表示など、一般にテキストの一部を他の部分よりも目立たせる手段を用いた表示をいう。

【0229】この例では、「半導体貿易」というトピックに適合する記事を検索するための検索条件として、「半導体、IC、調達などの単語を本文中に含む」という条件が設定されていたものとする。

【0230】記事は上記の条件を実際に満たしているもので、このことを明示するために本文一文目の「半導体」、「IC」、「調達」という単語が強調表示されている。

【0231】また、この変形例として、例えば「記事見出し」の行の「IC」という単語を強調表示してもよい。

【0232】このような強調表示により、ユーザは、表示中の記事がどのような根拠に基づいて検索され、提示されたのか理解することができる。

【0233】また、強調表示されている部分のテキストは、内容的に重要であることが多いので、ユーザは拾い読みにより効率的に記事内容を把握することが可能になると考えられる。

【0234】これは、例えば、レlevance・フィードバックのために提示された記事の有用性を判定する作業の効率化にもつながる。

【0235】図51、52、53に、適合した検索条件を記事中に強調表示することにより記事の有用性の判定が効率的になる例を示す。

【0236】図51(a)は、「自然言語処理」というトピックに適合する記事を検索するための検索条件の例である。

【0237】この例では、記事の本文中に「自然言語処理」、「NL」、「機械翻訳」、「かな漢字変換」という言語表現が出現するとその記事の得点が高くなる。

【0238】また、「自然言語」と「解析」という表現が同一の文中に出現すると、その記事の得点が高くな

る。

【0239】この他、記事を検索するための様々な条件が記されているものとする。

【0240】図51(b)は、図51(a)の検索条件を用いて検索され、ユーザに提示された記事の例である。この記事は「本文中に自然言語処理という言語表現を含む」という検索条件を満たしているので、記事中の「自然言語処理」という表現が強調表示されている。ここで、強調表示されている「自然言語処理」という表現を含む文の辺りを読んでみると、「このソフトウェアは、自然言語処理は用いずに、簡単な文字列マッチングにより検索を行う。」と書いてあるので、実際には自然言語処理に関する記事ではないことがすぐにわかる。

【0241】ユーザはこの時点でこの記事を読む必要がないと判断できるので、有用そうな記事のみを読んで情報収集を行ったり、効率的にレlevance・フィードバックを行ったりできる。

【0242】図52も、図51と同様に、記事が有用でないことを迅速に判断する例である。

【0243】この例での検索対象は英文テキストであり、「artificial intelligence (人工知能)」というトピックの検索条件が図52(a)である。

【0244】ここでは、「artificial」、「intelligence」などの単語を含む記事の得点が高くなるようにしている。

【0245】図52(b)は、図52(a)の検索条件を用いて検索されユーザに提示された記事の例であり、「artificial」という語が強調表示されている。図51と同様に、強調表示された語の周辺のみを拾い読みすることにより、この記事は「artificial hand (義手)」に関するものであって「artificial intelligence」とは無関係であるということが瞬時にしてわかる。

【0246】図51、52が有用でない記事の表示例であるのに対し、図53はユーザにとって有用である記事の表示例である。

【0247】図53(a)は、「パソコン新製品」というトピックに適合する記事を検索するための検索条件であり、記事に出現する単語として「ノートパソコン」、「ラップトップ」、「デスクトップ」などのパソコンの種類を表す表現や、「〇〇社」、「△△社」など、パソコンメーカーの名前などが指定されている。

【0248】図53(b)は、図53(a)のような検索条件により検索を行った結果得られ、ユーザに提示された記事の表示例である。

【0249】「△△社」が強調表示されているので、この記事で紹介されているパソコンのメーカーは〇〇社などのものではなく△△社のものであることがひと目でわかる。

【0250】同様に、「ノートパソコン」が強調表示されているので、発売したパソコンの種類はラップトップやデスクトップなどではなくノートパソコンであることがひと目でわかる。このように、提示されている記事がユーザにとって有用である場合にも、記事の内容を把握することが容易になると考えられる。

【0251】図50において、トピック検索条件と、そのうち記事が満足した検索条件をユーザに提示する例を示したが、これらの表示方法の例を説明する。

【0252】図54は、「半導体貿易」というトピックに適合する文書を検索するための検索条件の具体例である。

【0253】一行目の条件は、通常の文書検索で用いられるブール式の例であり、「半導体」「貿易」などの言語表現がANDやORなどの演算子で結合されている。

【0254】2行目の条件は、「半導体」と「貿易」という言語表現が同一文中に出現するという条件を表している。

【0255】また、例えば4行目の条件は、記事の記事見出し文字列の中に「半導体」、「メモリー」、「IC」などの言語表現が出現するという条件を表している。

【0256】図54のような検索条件により検索されユーザに提示された図50のような記事には、例えば図55のような情報を付加して表示する。

【0257】この例では、現在表示中の記事は「半導体貿易」というトピックに適合したものであることが明示されており、図54に示した「半導体貿易」のトピック検索条件がそのままユーザに提示されている。

【0258】そして、その下には、実際に記事が満足した条件が列挙されている。例えば、図50の第1文には「半導体」と「調達」という言語表現が出現しているので、図55の「記事1が満足した検索条件」のところには、満足された条件である「第1文：半導体（1回）、調達（1回）」が表示されている。

【0259】ここで、「（1回）」は出現回数を表す。

【0260】また、トピック検索条件の「単語：」という行に記された言語表現のうち、記事中に実際に出現したものは「半導体」、「IC」、「調達」の3つであったことが「記事1が満足した検索条件」の「単語：」という行に表示されている。

【0261】同時に、これらの出現した位置や、出現回数などの情報が表示されている。

【0262】さらに、トピック検索条件の1行目の「（半導体ORメモリー）AND（貿易OR調達）」というブール式は、図50の記事中に「半導体」と「調達」という表現が出現しているために満足されているので、「記事1が満足した検索条件」ではこのブール式が表示され、さらにその中の「半導体」と「調達」という表現が強調表示されている。

【0263】図56に、図55の変形例を示す。

【0264】図55がトピック検索条件と記事が実際に満足した検索条件とを別々に表示しているのに対し、図56ではトピック検索条件の中に記事が満足した検索条件を埋め込む形で表示している。

【0265】この例では、「半導体」や「調達」などの実際に満足された条件の語が強調表示されている。

【0266】これにより、トピック検索条件のうち何パーセントくらいの条件が記事によって満たされているか
10 がおおまかに把握できる。

【0267】以上のように、表示中のトピックの検索条件と、そのうち記事が実際に満足している検索条件の情報を提示することにより、その記事の有用性を判断しながら拾い読みをしたり、内容の把握を容易にすることが可能であると考えられる。

【0268】また、記事がどのような根拠から検索され提示されたのかがユーザにわかるようになるので、ユーザはよりきめ細かで効果的なレバンス・フィードバック用情報を情報フィルタリングサービス側に返すことが
20 可能になると考えられる。

【0269】（実施形態2の変形例1）次に、記事検索部24及び付加情報生成部26の他の構成例について説明する。

【0270】まず、記事検索部24は、変数iに1を代入した後（ステップS151）、i番目のトピック（トピック1）の検索条件をトピック記憶部22から取り出す（ステップS152）。この後、記事検索部24は、変数jに1を代入した後（ステップS153）、トピックi（トピック1）と到着記事j（到着記事1）との類似度を計算し、記事情報記憶部27に格納する（ステップS154）。この類似度計算は、通常の検索処理に相当するもので、記事情報記憶部18に格納されている記事の表現や検索インデックスが参照される。

【0271】ここで、実施形態2の図42との違いは、各記事が満足した検索条件に関する情報を必ずしも記事情報部27に記憶する必要がないという点だけである。

【0272】これは、実施形態2が、ユーザに提示する記事にその記事が何故検索されたかという情報を付加して提示するものであるのに対し、この変形例では、ユーザに提示する記事に、他のユーザがその記事をどのように
40 に読んでいたかという情報を付加して提示するものであるためである。

【0273】次いで、記事検索部24は、変数jの値を+1更新した後、そのときのjの値が到着記事数よりも大きいかなかを調べ（ステップS155、S156）、大きくない場合には類似度計算されてない記事が残っていると認識し、jの値が到着記事数よりも大きくなるまで、ステップS154～S156を繰り返す。到着した全ての記事に対してトピックiとの類似度の計算が終わると、記事検索部24は、到着した記事をユーザプロフ
50

ファイルとの類似度が高い順にソートし、記事のランキングを行う（ステップS157）。このランキングの結果は、記事情報記憶部27に記憶される。

【0274】この後、記事検索部24は、変数*i*の値を+1更新した後、そのときの*i*の値が全トピック数よりも大きいか否かを調べ（ステップS158、S159）、大きくない場合には類似度計算されてないトピックが残っていると認識し、*i*の値が全トピック数よりも大きくなるまで、ステップS152～S159を繰り返す。

【0275】図58に、付加情報生成部26の処理の流れを示す。

【0276】付加情報生成部26は、全てのユーザに対して以下の処理を行う。

【0277】まず、付加情報生成部26は、変数*i*に1を代入した後（ステップS161）、ユーザ*i*（ユーザ1）のユーザプロファイルをトピック記憶部22から取り出す（ステップS162）。次に、付加情報生成部26は、記事選出部25によって選出されたユーザ1に提示する記事と、これらの記事を受信する他のユーザに関する情報とを記事情報記憶部27から取り出す（ステップS163）。

【0278】この後、付加情報生成部26は、記事選出部25によって選出された記事に、これらの記事を受信する他のユーザに関する情報を付加してユーザ*i*に提示する（ステップS164）。そして、付加情報生成部26は、変数*i*の値を+1更新した後、そのときの*i*の値が全ユーザ数よりも大きいか否かを調べ（ステップS165、S166）、大きくない場合には付加情報が生成されてないユーザが残っていると認識し、*i*の値が全ユーザ数よりも大きくなるまで、ステップS162～S166を繰り返す。

【0279】例えば、図59のように、どのユーザにどの記事を送信するかという情報が記事選出部25により格納されていたとする。

【0280】この例では、例えばユーザ1には記事1、2を提示することが、ユーザ2には記事2、3、4を提示することが記されている。

【0281】付加情報生成部26は、ユーザ1に記事1を提示する際に、記事1を受信する他のユーザであるユーザ3、4に関する情報を付加して提示する。例えば記事1の受信人数を提示する場合、ユーザ1、3、4の3人という情報を付加したり、ユーザ1を除いた2人という情報を付加したりする。

【0282】同様に、ユーザ1に記事2を提示する際には、ユーザ2、ユーザ4に関する情報を付加して提示する。

【0283】図60に、あるユーザのために選出された記事の記事見出しのリストに、記事を受信した他のユーザに関する情報を付加してそのユーザに提示した表示例

を示す。

【0284】この例では、情報フィルタリングサービスの全ユーザ数は4、000人となっている。

【0285】そして、例えば記事番号1の記事を受信したユーザ数は250人だということがわかる。

【0286】図61に、あるユーザのために選出された記事の要約文または抜粋文に記事を受信した他のユーザに関する情報を付加して提示した表示例を示す。

【0287】図60と同様に受信ユーザ数の情報が示されている。

【0288】図62に、記事を受信した他のユーザに関する情報を記事本文のヘッダ情報として付加しユーザに提示した例を示す。

【0289】全4、000人のユーザ中、250人が表示中の記事を受信していることが明示されている。

【0290】図63に図62の変形例を示す。

【0291】この例では、記事の受信ユーザ数の内訳が表示されている。

【0292】表示中の記事を受信した250人のうち150人が男性、100人が女性であり、200人が日本人、30人が米国人、20人がその他の国のユーザであることがわかる。

【0293】さらに、250人のうち、180人は「半導体貿易」というトピックを選択しているユーザ、50人は「IC」というトピックを選択しているユーザ、20人は「半導体貿易」と「IC」の両方を選択しているユーザであることがわかる。この他、受信者の所属や年齢層など、プライバシー侵害にならない程度の統計情報を表示するようにしてもよい。

【0294】以上の説明のように、現在表示中の記事、他のどのようなユーザが何人受信しているかという情報がわかれば、その記事がどのくらい一般的な記事であるか、または特殊なユーザ層にだけ読まれる記事であるかがわかり、ユーザにとってどれくらい有用な記事であるかの判断材料となると考えられる。

【0295】例えば、図60のような記事情報を受信したユーザが、6つの記事全てを読んでいる時間がない場合、とりあえず一般常識的な情報だけを収集しようとして、記事番号4の記事のように多くのユーザが読んでいる記事だけを読むといった使い方が考えられる。

【0296】図64に、あるユーザや他のユーザが前回に行ったレバンス・フィードバック情報を今回提示する記事情報に付加して提示する表示例を示す。

【0297】この例では、今回到着した記事はb1～b4の4件であり、ユーザはこれらの全てあるいは一部に対して有用性の判定を行い、レバンス・フィードバックを行おうとしているものとする。

【0298】例えば、ユーザが記事b1に対して「有用ではない」という判定を行い、この情報を情報フィルタリングセンタ1側に送信すれば、情報フィルタリングセ

31

ンタ1側は記事b1のようなトピックの記事の優先度を下げるなどしてユーザプロフィールを修正し、次回からはユーザの要求に合致した記事がより多く提示されるようにすることが可能である。

【0299】図64では、この有用性の判定の参考情報として、ユーザが前回あるいはそれ以前に行った有用性判定に関する情報と、他のユーザの有用性判定に関する情報が提示されている。

【0300】この例では、ユーザが前回受信し、有用性判定を行った記事はa1～a6の6件あり、例えばユーザは記事a1に対しては「有用である」、記事a3に対しては「不要である」という判定をしたことがわかる。

【0301】一般に、人間による有用性の判定には一貫性がなく、同じような記事に対してもあるときは「有用である」とし、またあるときは「やや有用である」、というように違った判断をしてしまうことがあると考えられる。

【0302】このように一貫性に欠ける判定情報をフィードバックしてユーザプロフィールを修正しても、よりよいフィルタリングが行われるようになる保証はない。

【0303】この例のように、本人がこれまでにやってきた有用性判定情報にアクセスを許すことにより、今回の有用性の判定の信頼性や効率を高めることができると考えられる。また、ユーザの要求が時とともに変化した場合でも、自分の過去のフィードバック結果を参照しながら意識的に有用性判定の方針を変えようといった使いかたも考えられる。

【0304】また、図64では、本人の過去の判定情報に加えて、他のユーザの判定情報が表示されている。

【0305】例えば、記事a1は、他の250人のユーザによっても受信され、有用性の判定が行われており、このうち100人が「有用である」と判定し、100人が「やや有用である」と判定し、50人が「不要である」と判定したことがわかる。このように、他のユーザが過去に行った有用性判定の情報を参照して、自分のこれからの有用性判定の参考にしたり、また、自分が過去に行った有用性判定を直接訂正し、レlevance・フィードバックを再度実行してもらうなどが可能となる。

【0306】これによりより信頼性が高く効率のよいレlevance・フィードバックが行えると考えられる。

【0307】図64の変形例を図65に示す。

【0308】図64ではユーザが「有用」「やや有用」「不要」という離散的な評価値により有用性を判定しているのに対し、図65では連続的な得点により判定している。

【0309】「前回のrelevance feedback情報」の、他のユーザの判定情報のところには、他のユーザがつけた得点の平均値が表示されている。

【0310】例えば、今回の記事b1の内容が、前回の記事a1の内容に似たものである場合、ユーザは、前回

32

のa1に対する自分の判定が10点であったことから、今回のb1に対しても高得点を与えようといった使い方が考えられる。

【0311】また、前回の記事a5の行を見ると、自分は1点という低い評価を行ったのに対して、他のユーザの平均値は7.4点と比較的高い値になっている。

【0312】そこで、ユーザは、自分のa5に対する有用性の評価を撤回して、新たに評価値を付与しなおすといった使い方が考えられる。

10 【0313】このように、実施形態2のシステムでは、提示されている記事がユーザの選択したトピックのうちいずれに適合したものであるかなど、提示されている記事がいかなる検索条件を満足したのかがユーザに明示されるため、何故その記事が提示されているかがユーザに分かりやすくなり、記事の有用性の判定が容易になる。したがって、既にユーザに送信された各記事がユーザにとって有用であったか否かなどの情報についてユーザからフィードバックを受け、その情報を反映して検索条件を修正するレlevanceフィードバック機能をより有効利用できるようになる。

20 【0314】また、記事が選択された根拠の代わりに、提示されている記事が他のユーザによってどのように読まれているかをユーザに提示することによっても、他のユーザの判断を参考にしたレlevance・フィードバックが可能になり、レlevanceフィードバックの有効利用を図ることができる。

30 【0315】（実施形態3）次に、この発明の情報フィルタリングシステムの第3の実施形態について説明する。このシステム全体の構成は、図1と同様であり、ユーザ毎にユーザプロフィールを保持しており、そのユーザプロフィールを利用して記事の検索が行われる。ここで、ユーザプロフィールとは、前述したように、ユーザの関心が高いトピックに適合する記事を検索するための検索条件のことをいう。

40 【0316】図66には、実施形態3に係る情報フィルタリングセンタ1の構成が示されている。情報フィルタリングセンタ1は、図示のように、ユーザプロフィール生成部31、トピック記憶部32、記事情報抽出部33、記事検索部34、記事選出部35、要約・抄録生成部36、および記事情報記憶部37から構成されている。これら構成要素のうち、破線で囲まれているユーザプロフィール生成部31、記事情報抽出部33、記事検索部34、記事選出部35、要約・抄録生成部36は、例えば図1の中央処理装置14によって実行されるソフトウェアによって実現でき、またトピック記憶部32および記事情報記憶部37は記憶装置5によって実現できる。

50 【0317】ユーザプロフィール生成部31は、入力として個々のユーザの要求・興味を受けつける。ユーザの要求・興味は、「○○と××に関する記事が読みたい」

といった自然言語や、興味のあるトピックに頻出するキーワードの集合や、それらに優先順位や重みをつけたもの、または通常の文書検索における検索式のようなもので表される。

【0318】ユーザプロフィール生成部31は、これに対して単語抽出、同義語展開などの言語処理を行い、検索が可能となるような形式に変換してユーザプロフィールを作成する。ユーザプロフィールはユーザ毎にトピック記憶部32に記憶される。また、ユーザプロフィール生成部31は、既にユーザに送信された各記事がユーザにとって有用であったか否かなどの情報についてユーザからフィードバックを受け、その情報を反映してトピック記憶部32の検索条件を修正するというレlevanceフィードバック機能も有している。

【0319】図67に、キーワードとその重みで表現されたユーザプロフィールの例を示す。

【0320】この例では、ユーザが半導体に関する記事に興味をもっているために「メモリー」などの関連用語が羅列されており、それぞれの用語に類似度計算に利用する重みが定義されている。

【0321】記事情報抽出部33は、入力として情報源から到着した記事を受けつけ、これに対して形態素解析、構文解析、書式解析などを行い、記事の情報源や日付、文字や単語その他の文書構成要素の頻度情報や出現位置、5W1H的な情報などの抽出を行う。そして、記事をこれらの抽出された情報の集合体として表現する。例えば、出現した単語の頻度を要素とするベクトルにより記事を表現したり、5W1Hのテンプレートに実現値を代入したものにより表現したりする。記事情報抽出部33はまた、記事検索を高速に実現するためのインデクシング処理も行う。記事情報抽出部33により抽出された記事情報は、記事情報記憶部37に記憶される。

【0322】記事検索部34は、トピック記憶部32に記憶されている各トピックの検索条件と、記事情報抽出部33によって抽出された記事情報とを参照し、各トピックに適合する到着記事を検索する。これは、トピックと到着記事との類似度を算出することに相当する。この類似度は、検索の方式によって「トピックに適合する」「トピックに適合しない」などの離散値をとる場合もあるし、よく適合している記事ほど類似度の値が高くなるように連続値をとる場合もあるが、ここでは、より一般的である類似度が連続値をとる場合について説明する。この場合には、記事検索部34が各トピックについて行う処理は実施形態1、2と同様であり、まず、トピック記憶部32からトピックに適合する記事を検索するための検索条件を読み込む。次に、到着した記事の各々に対して、そのトピックとの類似度を計算する。この類似度計算は、通常の検索処理に相当するもので、記事情報記憶部に格納されている記事の表現や検索インデックスが参照される。記事の類似度、及びその記事が満足した検

索条件の情報は記事情報記憶部37に記憶される。到着した全ての記事に対して類似度の計算が終わると、すなわち、到着した全ての記事を検索対象とした検索処理が終わると、到着した記事はトピックとの類似度が高い順にソートされる。すなわち、記事のランキングが行われる。ランキングの結果も記事情報記憶部37に記憶される。

【0323】記事選出部35は、記事検索部34により記事情報記憶部37に格納された各トピックの検索結果の中から、各ユーザに提示する記事を選出する。例えば、「半導体貿易」、「低価格パソコン」、「人工知能」という3つのトピックを選択しているユーザに対しては、「半導体貿易」の検索結果、「低価格パソコン」の検索結果、「人工知能」の検索結果の3つを取り出し、これらのうち上位の記事の中からユーザに提示するものを選出する。

【0324】図68に、本実施形態3における要約・抄録生成部36の処理の流れを示す。要約・抄録生成部36は、各ユーザに対して以下の処理を行う。

【0325】まず、要約・抄録生成部36は、変数 i に1を代入し(ステップS171)、ユーザ i のユーザプロフィールをトピック記憶部32から取り出す(ステップS172)。次に、要約・抄録生成部36は、ユーザ i に提示する記事の集合と、各記事がユーザの選んだトピックのうちいずれに適合したかを示す情報とを記事情報記憶部37から取り出す。そして、要約・抄録生成部36は、変数 j に1を代入し、ユーザに提示する記事 j に対して、適合したトピックの情報を参照しながら、そのトピックに応じた長さの要約あるいは抄録を生成する(ステップS175)。

【0326】ここで、要約とは、記事の主題を端的に表すために原文をもとに生成したテキストをいい、抄録とは、重要文など、記事の原文テキストの一部をそのまま抜きだしたものをいう。

【0327】また、要約・抄録の「長さ」とは、原文に対する圧縮率、文数、段落数、文字数、あるいは提示するテキスト全体に占める割合などをいう。

【0328】本実施形態3で用いる要約・抄録生成の方法は、長さを2段階以上に調節できるものであれば、どのような方法であってもよい。

【0329】例えば、自然言語解析を用いた自動要約生成技術でもよいし、初めの一段落目のみ表示するかもしれない全文を表示するかという単純な方法でもよい。

【0330】次に、要約・抄録生成部36は、現在の j の値がユーザ i に提示する記事数よりも大きいか否かを調べ(ステップS176)、大きくないならば、要約・抄録生成処理を行っていない記事が残っていると判断し、 j の値がユーザ i に提示する記事数よりも大きくなるまでステップS175、S176を繰り返す。

【0331】この後、要約・抄録生成部36は、ユーザ

iに該当記事の要約または抄録を提示し(ステップS177)、その後、現在のiの値が全ユーザ数よりも大きいか否かを調べる(ステップS178)。現在のiの値が全ユーザ数よりも大きくないならば、大きくなるまでステップS172~S178の処理が繰り返される。次に、記事の適合したトピックに応じた長さの要約・抄録を生成する手順を図を用いて説明する。

【0332】図69は、あるユーザが選択したトピックと、それらの間の優先度の例を示す。

【0333】この例では、ユーザはトピックA、B、C、Dの4トピックを選択しており、これらに関する記事を求めている。また、トピックA、B、C、Dの順で優先度が高くなっている。

【0334】この優先度は、情報フィルタリングサービスセンタ1側が設定してもよいし、ユーザが指定してもよい。ここでは、ユーザが指定したものであるとすると、このユーザは、例えばトピックBに適合する記事よりもトピックAに適合する記事により興味があるということになる。

【0335】図70に、図69のトピックを選択しているユーザに提示する記事のリストとそれらに適合したトピックの例を示す。

【0336】この例では、ユーザのために記事1~4の4つの記事が選出されている。記事1、2はトピックAに、記事3はトピックBに、記事4はトピックCおよびDに適合したものである。

【0337】図71に、図70のような場合にユーザに提示する記事情報の概念図を示す。記事1、2は、ユーザの選んだトピックのうちもっとも優先度が高いトピックAに適合したものであるため、比較的長い要約あるいは抄録が提示されている。これに対して、記事4は、ユーザの選んだトピックのうちもっとも優先度が低いトピックC、Dに適合したものであるため、非常に短い要約あるいは抄録が提示されている。

【0338】このように、トピックの優先度に応じて段階的に抄録の長さを変化させる。

【0339】なお、図71では、要約あるいは抄録の長さを図中の面積であらわしているが、優先度の最も高いトピックに適合した記事の要約・抄録が必ずしも最も長くなるとは限らない。

【0340】例えば、抄録の長さとして、原文の文数をもとにした圧縮率を採用するとし、記事1の原文が5文、記事4の原文が20文であったとする。

【0341】このとき、記事1は優先度の高いトピックAの記事なので圧縮率100%とし、記事4は優先度の低いトピックC、Dの記事なので圧縮率50%として表示したとすると、記事1の抄録は原文のままなので5文となり、一方、記事4の抄録は10文となる。

【0342】以上のような機能により、ユーザはトピック毎に異なる詳細度の記事を読むことができる。

【0343】ユーザが選択したトピック間に明確な優先度がある場合には有効であると考えられる。

【0344】(レlevance・フィードバック) 文書検索における技術に、検索結果の文書に対してユーザに有用性の判定を行ってもらい、その結果を利用して検索式中の単語の重み値を変更することにより、ユーザの求めるものにより近い文書を検索するレlevance・フィードバックがある。

【0345】情報フィルタリングの分野でもこの機能は実現されつつある。

【0346】本実施例においては、レlevance・フィードバックの際に得られる有用性判定情報を、要約・抄録の長さに反映させることが可能である。

【0347】例えば、既に説明した図71のような情報提示に対して、ユーザが「記事3が非常に有用であった」という情報を返したとする。

【0348】同時に、図70で示したトピック間の優先度が具体的には図72のように重要度の値の大小で定義されていたとする。

【0349】このとき、ユーザによって特に有用であるとされた記事3は、トピックBに適合した記事であるので、何らかの計算によりトピックBの重要度の値をより大きくし、今度はトピックBに適合する記事の長さをより長くして提示することが有用である。

【0350】図73に、このようなフィードバックを行った場合に、次回のフィルタリングでユーザに提示される記事情報の例を示す。

【0351】図71ではトピックAの優先度が最も高かったが、この図ではフィードバックによりトピックBの優先度がもっとも高くなっており、トピックBに適合した記事1'の要約あるいは抄録が最も長くなっている。

【0352】(実施形態3の変形例1) 次に、要約・抄録生成部36の他の構成例について説明する。

【0353】図74に、本変形例における要約・抄録生成部36の処理の流れを示す。

【0354】要約・抄録生成部36は、各ユーザに対して以下の処理を行う。

【0355】まず、ユーザiのユーザプロファイルをトピック記憶部32から取り出す(ステップS181、S182)。次に、ユーザiに提示する記事の集合と、各記事に予め付加されている日付、新聞社、朝刊/夕刊、見出しの大きさや行数、何面に掲載されたか、などの属性情報とを記事情報記憶部37から取り出す(ステップS183)。そして、ユーザiに提示する各記事に対して、属性情報を参照しながら、それに応じた長さの要約あるいは抄録を生成する(ステップS185)。ここで、「要約」、「抄録」、および「長さ」の意味は、実施形態3と同様である。また、以下の処理も実施形態3と同様である。

【0356】以下に、記事の属性に応じた長さの要約・

抄録を生成する手順を図を用いて説明する。

【0357】図75に、あるユーザに提示するために記事選出部35により選出された記事の例を示す。この例では、記事に予め付加されている属性として、発行日などの日付の情報を採用している。

【0358】記事1~4の日付は、それぞれ5月26日、23日、23日、20日となっている。

【0359】例えば一週間毎に情報をまとめて配信するような形態のサービスでは、このように、記事には新しいものと古いものが混在している可能性がある。

【0360】図76に、図75のような場合にユーザに提示する記事情報の概念図を示す。この例では、新しい記事ほど要約あるいは抄録の長さを長くして表示している。

【0361】例えば、5月26日付けの記事1は詳しく表示されているが、5月20日付けの記事4は簡単に表示されている。

【0362】同様に、記事が情報フィルタリングセンターに到着した時間や朝刊/夕刊の情報などに応じて要約あるいは抄録の長さを変えてもよい。

【0363】また、時間的属性として曜日を採用し、例えば「月曜日の記事は他の曜日の記事よりも詳しく表示する」などといった処理も可能である。

【0364】図77に、属性として新聞社が採用されている場合の、あるユーザに提示するために選出された記事の例を示す。

【0365】この例では、記事1は〇〇新聞から到着したもの、記事2、3は△△新聞から到着したもの、記事4は××新聞から到着したものである。

【0366】ここで、ユーザあるいは情報フィルタリングサービス側が、〇〇新聞、△△新聞、××新聞の順で優先度を定めていたとすると、ユーザには例えば図78のような情報が提示される。

【0367】記事1はもっとも優先度の高い〇〇新聞の記事なので、長い要約あるいは抄録が提示されており、一方記事4はもっとも優先度の低い××新聞の記事なので、短い要約あるいは抄録が提示されている。

【0368】ここでは新聞社で説明したが、何面の記事か、位置、社会面などの分類など、各記事の発信者によ

って予め付与されている様々な属性に応じて要約あるいは抄録の長さを変化させることが可能である。

【0369】(レバンス・フィードバック) 本変形例においても実施形態3と同様に、レバンス・フィードバックの際に得られる有用性判定情報を、要約・抄録の長さに反映させることが可能である。例えば、既に説明した図78のような情報提示に対して、ユーザが「記事2、3が非常に有用であった」という情報を返したとする。

10 【0370】記事2、3は共に「△△新聞」という属性をもつ記事であるので、何らかの計算により△△新聞の重要度の値をより大きくし、今度は△△新聞に適合する記事の長さをより長くして提示することが有用であるかも知れない。

【0371】図79に、このようなフィードバックを行った場合に、次のフィルタリングでユーザに提示される記事情報の例を示す。

20 【0372】図78では〇〇新聞の優先度が最も高かったが、この図ではフィードバックにより△△新聞の優先度がもっとも高くなっており、△△新聞という属性をもつ記事1'の要約あるいは抄録が最も長くなっている。

【0373】以上のように、実施形態3のシステムでは、記事の種類(その記事が満足したトピックなどの検索条件、またはその記事の発行日時などの記事そのものの属性)に応じた長さの要約あるいは抄録が作成されてそれがユーザに提示されるため、ユーザに提示されるテキストのうち、ユーザにとって有用であるテキスト情報の占める割合が高くなる。これにより、効率的な情報収集が可能になる。

30 【0374】(実施形態4) 次に、この発明の第4の実施形態に係る情報フィルタリングシステムを説明する。全体的なシステム構成は実施形態1と同じであるので、ここでは、実施形態1との差異について説明する。

【0375】図3の記事間類似度計算部16は図14で示したような記事間類似度計算処理を行うが、ある記事iと記事jとの記事間類似度計算には、次の計算式が使用される。

【0376】

【数1】

$$S_{ij} = |C_i \cap C_j| / (|C_i \cup C_j|) \quad \dots (1)$$

ここで、 C_i は記事i中に含まれる異なる単語の集合、 C_j は記事j中に含まれる異なる単語の集合であり、 $|A|$ は集合Aの要素数を取ることを意味し、 \cap は集合間の共通要素の集合を、 \cup は合併集合を取ることを意味する。この場合、記事iと記事jの間の類似度 S_{ij} は、記事間で重複する単語の比率ということになる。

39

40

類似度計算式の変形例として、例えば以下の式があげられる。

*【0377】
*【数2】

$$S_{ij} = |C_i \cap C_j| \quad \dots (2)$$

$$S_{ij} = |C_i \cap C_j| / \sqrt{|C_i| \cdot |C_j|} \quad \dots (3)$$

$$S_{ij} = |C_i \cap C_j| / (|C_i| + |C_j| - |C_i \cap C_j|) \quad \dots (4)$$

$$S_{ij} = 2 \cdot (x_i \cdot x_j) / (x_i \cdot x_i + x_j \cdot x_j) \quad \dots (5)$$

$$S_{ij} = x_i \cdot x_j \quad \dots (6)$$

$$S_{ij} = x_i \cdot x_j / \sqrt{(x_i \cdot x_i) (x_j \cdot x_j)} \quad \dots (7)$$

$$S_{ij} = (x_i \cdot x_j) / (x_i \cdot x_i + x_j \cdot x_j - x_i \cdot x_j) \quad \dots (8)$$

上記の類似度計算式の変形例の中で、 x_i と x_j はそれぞれ記事 i と記事 j に含まれる単語の頻度ベクトルである。

【0378】上記の類似度計算では、記事中の全単語を対象にしているが、これを数種の品詞の単語に限定することも可能である。例えば、品詞を名詞と動詞だけに限定して類似度を計算することにしても構わない。

※【0379】また、記事間の類似度計算においては、見出しや一文目など書式上のフィールドごとに類似度を求め、その類似度の荷重平均を全体の記事間類似度として定義することもできる。この場合、式1に対応する類似度としては、以下の通りになる。

※【0380】

※【数3】

$$S_{ij} = \sum_f W_f \cdot S_{fij} \quad \dots (9)$$

$$S_{fij} = |C_{fi} \cap C_{fj}| / (|C_{fi} \cup C_{fj}|) \quad \dots (10)$$

ここで、 C_{fi} は記事 i のフィールド f に含まれる単語の集合、 C_{fj} は記事 j のフィールド f に含まれる単語の集合である。

【0381】フィールドとしては、見出しや一文目、一段落目など、文書の先頭文字の空白やインデントの情報、句点などの存在によって検出することができる。式2から8についても同様の変形が可能である。

【0382】また、上記のような数値的な類似度計算を情報をした後、構文情報などのチェックの過程を設け、ある閾値以上の類似度を持つ記事間であっても、類似記事としないように変形が可能である。例えば、新聞記事においては、一文目の主語（具体的には助詞「は」に前接している固有名詞）が重要な役割を果たす。この主語が記事同士異なる場合、類似記事としないようにする。

【0383】次に、図80を参照して、図3の提示情報生成部17による提示情報生成処理について説明する。

【0384】実施形態1では、重複記事の提示を回避するために、重複記事集合から、代表としてユーザに提示する記事を選択する場合について説明したが、ここでは、関連記事同士がグループ化あるいは関連づけされて

30 ユーザに提示される。

【0385】すなわち、まず、記事情報記憶部18から記事選出部15により選択された記事の情報が読み込まれる（ステップS201）。次いで、上述の式を用いることにより記事間類似度計算部16によって、選択された記事間の類似度計算が行われ、互いに類似度の高い記事の集合が求められる（ステップS202）。そして、関連記事同士のグループ化、関連づけ、あるいは特定の記事の選択といった出力制御が行われ、それがユーザに提示される（ステップS203）。

40 【0386】ここで、グループ化とは、関連記事同士が並んでユーザに提示されるように記事の出力リストを揃えることを意味する。また、関連づけとしては、例えば、ある記事とそれに関連する記事とを連結するリンク情報を用いてハイパーテキストを生成し、それをユーザに提示することなどがあげられる。また、特定記事選択では、関連記事の中から1つまたは幾つか記事が選択し、選択された記事だけがユーザに提示される。

50 【0387】この様なグループ化、関連づけなどを行うことにより、関連のあるテキスト記事が順不同でバラバラにユーザに出力されるといった事態を防止できる。よ

41

って、利用者は、関連記事を効率よく整理して読むことが可能となる。

【0388】(実施形態5)次に、この発明の第5の実施形態に係る情報フィルタリングシステムについて、実施形態1との差異を中心に説明する。実施形態5の構成を図81に示す。実施形態との相違は、利用者へ出力した記事を格納する送付記事記憶部19を有することにある。

【0389】送付記事記憶部19には、記事を提供した日付情報と共に利用者へ提供された記事が、利用者とは対応づけて記憶される。これは、利用者への記事の提供時に行われるものである。

【0390】図82に提示情報生成部の処理の流れを示す。まず始めに、記事選択部15により選出された記事の情報を読み込む(ステップS211)。そして、記事情報記憶部18に格納されている選出された当日記事と送付記事記憶部19に格納されている前日以前の記事とを参照し、記事間類似度計算部16により前日以前の記事をも対象とした記事間類似度計算が行われて、重複記事集合が求められる(ステップS212)。

【0391】この場合、重複記事集合 ϕ_k は、ある記事jを核にして、以下のように定義することができる。

【0392】

【数4】

$$\phi_k \text{ where } S_{ij} > T \text{ for } \forall \text{記事 } i \in \phi_k$$

すなわち、具体的には、記事選出部15により選出された記事を上位からスキャンし、その記事jに対して類似度が一定の閾値以上に入る記事を求め、その記事を重複記事とし重複記事集合を求めるものである。

【0393】この後、前述したような関連する記事同士のグループ化、関連づけ、あるいは特定の記事の選択といった出力制御が行われ、それがユーザにフィルタリング結果として提示される(ステップS213)。

【0394】図83には、ユーザに出力するフィルタリング結果の出力処理の流れが示されている。選出されている上位の記事から順に、重複記事集合を持つか否かを判定し(ステップS221、S222)、重複記事を持たない場合は、その記事(例えばタイトルと新聞社の情報など)を出力する(ステップS223)。一方、重複記事を持つ場合、その重複記事集合が当日記事のみからなるか否かを調べ(ステップS224)、当日記事のみからなる場合はマーク2を、それ以前の記事も含む場合はマーク1を出力するとともに重複記事集合を出力する(ステップS225、S226、S227)。選出されている残りの記事についても、ステップS222～S227の処理が同様に行われる(ステップS228、S229)。重複記事集合を出力する際には、タイトルをフラットなテキストとして出力する場合は、重複記事を並

42

べて出力することになる(グループ化)。図84はその出力例である。直線で囲まれている記事同士が、重複記事であることを意味している。また、□は当日の記事のみからなる重複記事集合であり、△はそれ以前の記事を含むことを意味するマークである。「8/4」などは記事の日付である。一方、ハイパーテキストとして出力できる場合、代表記事だけを一番上の階層に表示し、その代表記事に他の重複記事を関連づけておくことができる。このハイパーテキストの表示例を図85～図87に示す。

【0395】図85において、□は当日の重複記事集合を、△はそれ以前の重複記事集合を持ったことを意味しており、それぞれその代表記事のタイトルが表示されている。図85に示した最上位階層の記事リストを出力する際に、それぞれのマークには、図86や図87に示す重複記事集合の情報へのリンク情報を付与しておくことになる。これは、HTML(Hyper Text Markup Language)の記法を用いること等、公知の技術によって実現可能である。この場合、図85のマーク□を利用者が画面上で選択すると、図86の重複記事情報が表示され、また図85のマーク△を利用者が画面上で選択すると、図87の重複記事情報が表示されることになる。

【0396】このように、当日の記事だけからなる記事グループか、それ以前の日の記事も含まれているのかを区別するための情報を付加して送ることにより、利用者は、関連記事をさらに効率よく整理して読むことが可能となる。

【0397】なお、以上の実施形態1～5およびそれらの変形例は必要に応じて種々組み合わせる利用することができる。また、以上の説明は、通信網を介してフィルタリング結果をセンタからユーザに送るネットワークシステムとして実現した例について述べたが、この発明の情報フィルタリングシステムの中核をなすセンタの機能は個人の計算機上で構築することもできる。この場合、ユーザ端末とフィルタリングシステムが一体となった形態となり、ユーザ端末とフィルタリングシステム間には通信網は存在しないことになる。

【0398】

【発明の効果】以上説明したように、本発明によれば、情報フィルタリングによってユーザに提示される記事同士の関連性がユーザにとって明確になり、記事内容の理解が容易になる。特に時間の経過と共に状況が変わっていくような出来事についてその経緯を把握することや、連載記事など複数の記事にわたる情報を把握することが容易になり、フィルタリングシステムの性能向上を図ることができる。また、複数の情報源から得られた同一内容に関する記事が、重複してユーザに提示されることが自動的に回避できる。

【0399】また、提示されている記事がユーザの選択

したトピックのうちいずれに適合したものであるかがユーザに明示されるため、ユーザが記事内容を理解することが容易になる。また、提示されている記事が他のユーザによってどのように読まれているかが明示されるため、ユーザは一般的に読まれている記事や一部のユーザに読まれている記事などを識別することができる。さらに、過去に提示された記事に対してユーザが行った有用性の判定、及び他のユーザが行った有用性の判定情報へのユーザのアクセスを許すことにより、一貫性を保ったレlevance・フィードバックや他のユーザの判断を参考にしたレlevance・フィードバック、さらには過去に行った有用性判定を修正してレlevance・フィードバックをかけ直すことが可能となる。

【0400】また、さらに、記事に適合したトピックに応じた長さの要約あるいは抄録や、記事の持つ属性に応じた長さの要約あるいは抄録がユーザに提示されるため、ユーザに提示されるテキストのうち、ユーザにとって有用であるテキスト情報の占める割合を高くでき、効率的な情報収集が可能になる。

【0401】また、関連する記事同士がグループ化あるいは関連づけられて利用者に提供されるため、利用者の手間を大幅に軽減することができる。さらに、当日配信された記事間のみでなく、前日以前にユーザに出力した記事との間の類似度も求め、出力記事には、当日の記事だけからなる記事グループか、それ以前の日の記事も含まれているのかを区別するための情報を付加することにより、利用者は、関連記事をさらに効率よく整理して読むことが可能となる。

【図面の簡単な説明】

【図1】この発明の各実施形態が適用される情報フィルタリングシステム全体のシステム構成を示すブロック図。

【図2】図1の情報フィルタリングシステムの運用形態を概念的に示す図。

【図3】この発明の第1の実施形態に係わる情報フィルタリングシステムに設けられる情報フィルタリングセンタの構成を示すブロック図。

【図4】同第1実施形態のシステムにおけるユーザプロフィール生成処理の流れを示すフローチャート。

【図5】同第1実施形態のシステムにおける記事情報抽出処理の流れを示すフローチャート。

【図6】同第1実施形態のシステムにおける記事の表現例を示す図。

【図7】同第1実施形態のシステムにおける記事の他の表現例を示す図。

【図8】同第1実施形態のシステムにおける記事検索処理の流れを示すフローチャート。

【図9】同第1実施形態のシステムにおいて記事検索処理によってランキングされた到着記事の様子を示す図。

【図10】同第1実施形態のシステムにおける記事選出

処理の流れを示すフローチャート。

【図11】同第1実施形態のシステムにおいて図9のようなランキング結果が得られている場合にその上位10件を選出した例を示す図。

【図12】同第1実施形態のシステムにおいて図8のようなランキング結果が得られている場合にユーザプロフィールとの類似度が0.86以上の記事を選出した例を示す図。

【図13】同第1実施形態のシステムにおいて一人のユーザに対して複数の検索及びランキングが行われる場合にこれら複数のランキング結果の上位部分をマージしてユーザに提示する記事を選出する様子を示す図。

【図14】同第1実施形態のシステムにおける記事間類似度計算処理の流れを示すフローチャート。

【図15】同第1実施形態のシステムにおいて異なる情報源から到着した記事の例を示す図。

【図16】同第1実施形態のシステムにおける提示情報生成処理の流れを示すフローチャート。

【図17】同第1実施形態のシステムにおいて一件のプレスリリースから重複記事が派生する様子を示す図。

【図18】同第1実施形態のシステムにおいて一件の出来事から重複記事が作成される様子を示す図。

【図19】同第1実施形態のシステムにおいて図15の4つの記事に対して記事間類似度計算を行った結果得られる重複記事集合の例を示す図。

【図20】同第1実施形態のシステムにおいて排除された重複記事に関する情報を記事の本文情報に付加して提示する例を示す図。

【図21】同第1実施形態のシステムにおける関連記事情報の表示形態を示す図。

【図22】同第1実施形態のシステムにおける関連記事情報の他の表示形態を示す図。

【図23】同第1実施形態のシステムにおける関連記事情報のさらに他の表示形態を示す図。

【図24】同第1実施形態のシステムにおける関連記事情報の表示画面切り替え処理の流れを示すフローチャート。

【図25】同第1実施形態のシステムにおける関連記事情報の他の表示画面切り替え処理の流れを示すフローチャート。

【図26】同第1実施形態のシステムにおいて図20のような記事の重複が起こっている場合にユーザに提示する記事の一覧表を重複記事情報とともに表示した例を示す図。

【図27】同第1実施形態のシステムにおける記事間類似度計算処理の流れを示すフローチャート。

【図28】同第1実施形態のシステムにおいて今回記事選出部により選出された記事の集合と前回ユーザに提示された記事の集合の例を示す図。

【図29】同第1実施形態のシステムにおける提示情報

生成処理の流れを示すフローチャート。

【図30】同第1実施形態のシステムにおいて今回の記事の本文情報に前回までの関連記事の情報を付加して提示する例を示す図。

【図31】同第1実施形態のシステムにおいて今回の記事の本文情報に前回までの関連記事の情報を付加して提示する他の例を示す図。

【図32】同第1実施形態のシステムにおいて今回の記事の本文情報中に前回までの関連記事の情報を埋め込んで提示する例を示す図。

【図33】同第1実施形態のシステムにおいて図32の第一文が選択された場合にその文と関係の深い前回までの記事のリストが表示される様子を示す図。

【図34】同第1実施形態のシステムにおいて図33の「〇〇沖で地震 マグニチュード4」という関連記事を選択した場合にその記事の本文を表示した例を示す図。

【図35】同第1実施形態のシステムにおける記事間類似度計算処理の流れの他の例を示すフローチャート。

【図36】同第1実施形態のシステムにおける提示情報生成処理の流れの他の例を示す図。

【図37】同第1実施形態のシステムにおいて今回の記事の本文情報を今回の他の関連記事の情報とともに提示する他の例を示す図。

【図38】同第1実施形態のシステムにおいて今回の記事の本文情報を今回の他の関連記事の情報とともに提示する他の例を示す図。

【図39】同第1実施形態のシステムにおいて記事間類似度を記事の提示順序に反映させる例を示す図。

【図40】この発明の第2の実施形態に係わる情報フィルタリングシステムで使用されるユーザプロフィールを概念的を示す図。

【図41】同第2実施形態のシステムに設けられる情報フィルタリングセンタの構成を示すブロック図。

【図42】同第2実施形態のシステムにおける記事検索処理の流れを示すフローチャート。

【図43】同第2実施形態のシステムにおいてランキングされた到着記事を概念的に示す図。

【図44】同第2実施形態のシステムにおける記事選出処理の流れを示すフローチャート。

【図45】同第2実施形態のシステムにおけるトピックとその検索結果の例を示す図。

【図46】同第2実施形態のシステムにおける付加情報生成処理の流れを示すフローチャート。

【図47】同第2実施形態のシステムにおいてユーザのために選出された記事の記事見出しのリストに各記事が適合したトピックの情報を付加してそのユーザに提示する様子を示す図。

【図48】同第2実施形態のシステムにおいてユーザに対して各トピックに適合した記事の件数情報を提示する様子を示す図。

【図49】同第2実施形態のシステムにおいてユーザのために選出された記事の要約文・抜粋文あるいは本文をトピック別にまとめてそのユーザに提示する様子を示す図。

【図50】同第2実施形態のシステムにおいて記事が満足した検索条件に関する情報を記事本文のヘッダ情報として付加してユーザに提示する様子を示す図。

【図51】同第2実施形態のシステムにおいて適合した検索条件を記事中に強調表示する様子を示す図。

10 【図52】同第2実施形態のシステムにおいて適合した検索条件を記事中に強調表示する他の列を示す図。

【図53】同第2実施形態のシステムにおいて適合した検索条件を記事中に強調表示するさらに他の列を示す図。

【図54】同第2実施形態のシステムにおいてあるトピックに適合する文書を検索するための検索条件の具体例を示す図。

【図55】同第2実施形態のシステムにおいて図54の検索条件により検索されユーザに提示された記事に対して付加する検索条件の表示例を示す図。

20 【図56】同第2実施形態のシステムにおいて図54の検索条件により検索されユーザに提示された記事に対して付加する検索条件の他の表示例を示す図。

【図57】同第2実施形態のシステムにおける記事検索処理の他の例を示すフローチャート。

【図58】同第2実施形態のシステムにおける付加情報生成処理の他の例を示すフローチャート。

【図59】同第2実施形態のシステムにおいて複数のユーザそれぞれとそれらユーザに送信する記事との関係を示す図。

【図60】同第2実施形態のシステムにおいてあるユーザに対して選出された記事の記事見出しのリストに、記事を受信した他のユーザに関する情報を付加して提示する様子を示す図。

【図61】同第2実施形態のシステムにおいてあるユーザに対して選出された記事の要約文または抜粋文に記事を受信した他のユーザに関する情報を付加して提示する様子を示す図。

【図62】同第2実施形態のシステムにおいて記事を受信した他のユーザに関する情報を記事本文のヘッダ情報として付加しユーザに提示する様子を示す図。

【図63】同第2実施形態のシステムにおいて記事を受信した他のユーザに関する情報を記事本文のヘッダ情報として付加しユーザに提示する他の例を示す図。

【図64】同第2実施形態のシステムにおいてあるユーザや他のユーザが前回に行ったレlevance・フィードバック情報を今回提示する記事情報に付加して提示する表示例を示す図。

50 【図65】同第2実施形態のシステムにおいてあるユーザや他のユーザが前回に行ったレlevance・フィードバ

ック情報を今回提示する記事情報に付加して提示する他の表示例を示す図。

【図66】この発明の第3の実施形態に係わる情報フィルタリングシステムに設けられる情報フィルタリングセンタの構成を示すブロック図。

【図67】同第3実施形態のシステムにおけるキーワードとその重みで表現されたユーザプロファイルの例を示す図。

【図68】同第3実施形態のシステムにおける要約・抄録生成処理の流れを示すフローチャート。

【図69】同第3実施形態のシステムにおいてユーザが選択したトピックとそれらの間の優先度の例を示す図。

【図70】同第3実施形態のシステムにおいて図69のトピックを選択しているユーザに提示する記事のリストとそれらに適合したトピックの例を示す図。

【図71】同第3実施形態のシステムにおいてユーザに提示される記事情報を概念的に示す図。

【図72】同第3実施形態のシステムにおいてユーザが選択したトピックとそれらの間の優先度の例を示す図。

【図73】同第3実施形態のシステムにおいてフィードバックを行った場合に次回のフィルタリングでユーザに提示される記事情報の例を示す図。

【図74】同第3実施形態のシステムにおける要約・抄録生成処理の流れの他の例を示すフローチャート。

【図75】同第3実施形態のシステムにおいて記事選出部により選出された記事の例を示す図。

【図76】同第3実施形態のシステムにおいてユーザに提示される記事情報の他の例を概念的に示す図。

【図77】同第3実施形態のシステムにおいて属性として新聞社が採用されている場合にあるユーザに提示するために選出された記事の例を示す図。

【図78】同第3実施形態のシステムにおいて図77の

【図7】

記事1の表現:

情報源 : ○○新聞 ××月××日朝刊 2面
 文字数 : ○○文字
 記事見出し : 「○○社 ○○メモリ増産体制」
 トピック : 半導体
 日時 : △△月△△日
 場所 : ○○県○○工場
 主語 : (株) □□
 主動詞 : 増産した
 ……

場合にユーザに提示される記事情報を概念的に示す図。

【図79】同第3実施形態のシステムにおいてフィードバックを行った場合に次回のフィルタリングでユーザに提示される記事情報の他の例を示す図。

【図80】この発明の第4の実施形態に係わる情報フィルタリングシステムにおける提示情報生成処理の流れを示すフローチャート。

【図81】この発明の第5の実施形態に係わる情報フィルタリングシステムに設けられた情報フィルタリングセンタの構成を示すブロック図。

【図82】同第5実施形態のシステムにおける提示情報生成処理の流れを示すフローチャート。

【図83】同第5実施形態のシステムにおける重複記事集合の出力処理の流れを示すフローチャート。

【図84】同第5実施形態のシステムにおける利用者への記事提示例を示す図。

【図85】同第5実施形態のシステムにおけるハイパーテキストによる利用者への記事提示例を示す図。

【図86】同第5実施形態のシステムにおけるハイパーテキストによる利用者への記事提示例を示す図。

【図87】同第5実施形態のシステムにおけるハイパーテキストによる利用者への記事提示例を示す図。

【符号の説明】

1…情報フィルタリングセンタ、2…情報源、3…ユーザ端末、10…ユーザプロファイル、11、21、31…ユーザプロファイル生成部、12…ユーザプロファイル記憶部、13、23、33…記事情報抽出部、14、24、34…記事検索部、15、25、35…記事選出部、16…記事間類似度計算部、17…提示情報生成部、19…送付記事記憶部、22、32…トピック記憶部、26…付加情報生成部、36…要約・抄録生成部。

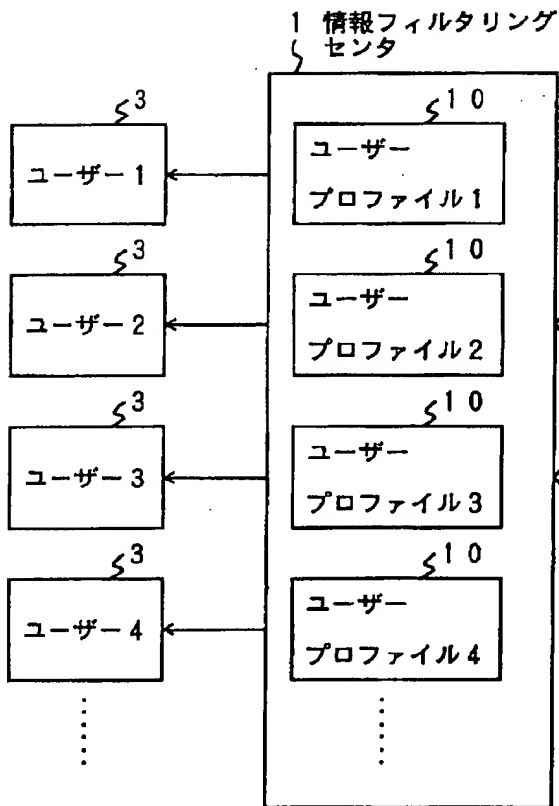
【図11】

順位	到着記事名	ユーザープロファイルとの類似度
1		0.97
2		0.96
3		0.96
4		0.94
5		0.91
6		0.89
7		0.89
8		0.87
9		0.85
10		0.79

10件

Figure 1 is a block diagram of a text information processing system. The system includes multiple user terminals (3) connected to a central information filtering center (1) via a communication network (通信網). Each user terminal (3) contains a text information receiving unit (8) and a text information display unit (9). The central information filtering center (1) contains a text information receiving unit (6), a central processing unit (4), a memory unit (5), and a text information transmitting unit (7). Multiple information sources (2) are connected to the text information receiving unit (6) of the central information filtering center (1).

【図2】



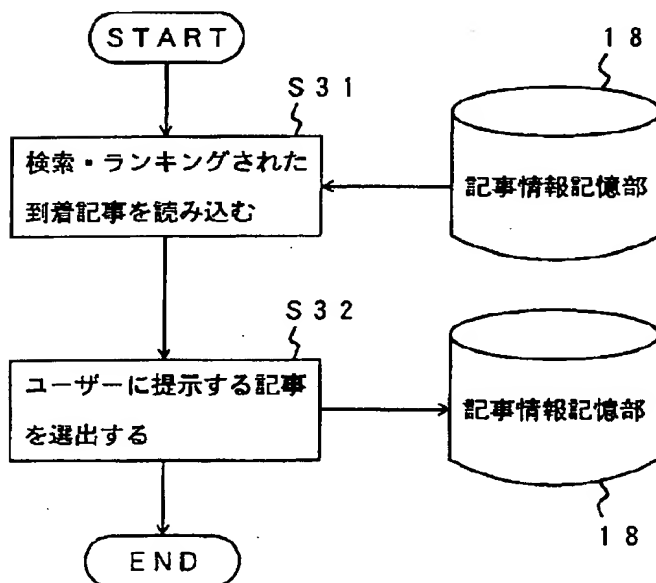
【図67】

キーワード	重み
DRAM	5
SRAM	5
EPROM	5
メモリー	7
集積回路	7
半導体	10
半導体ユーザ協議会	3

【図69】

ユーザー1が選択しているトピック 優先度	
トピックA	1
トピックB	2
トピックC	3
トピックD	4

【図10】



【図12】

順位	到着記事名	ユーザープロフィールとの類似度
1		0.97
2		0.96
3		0.96
4		0.94
5		0.91
6		0.89
7		0.89
8		0.87

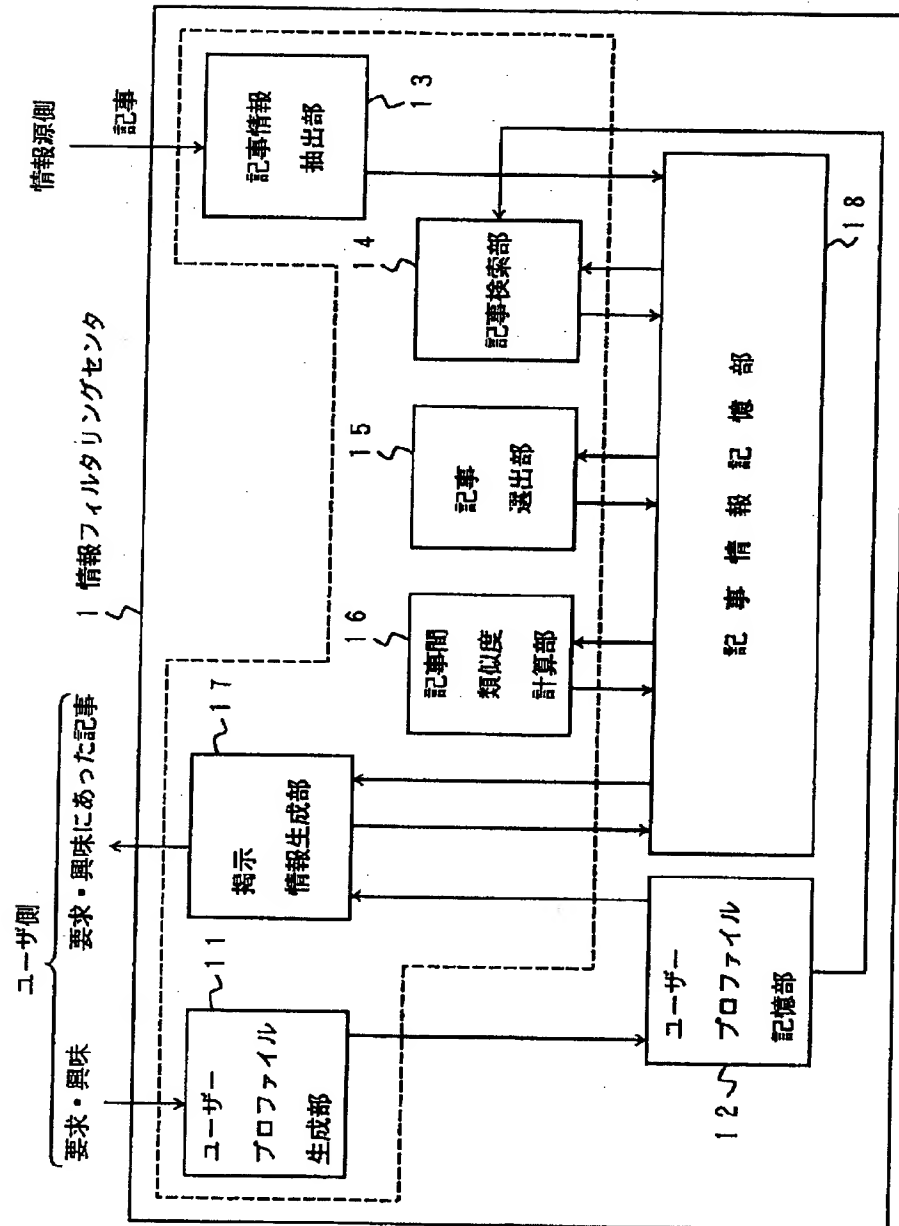
0.86以上

【図15】

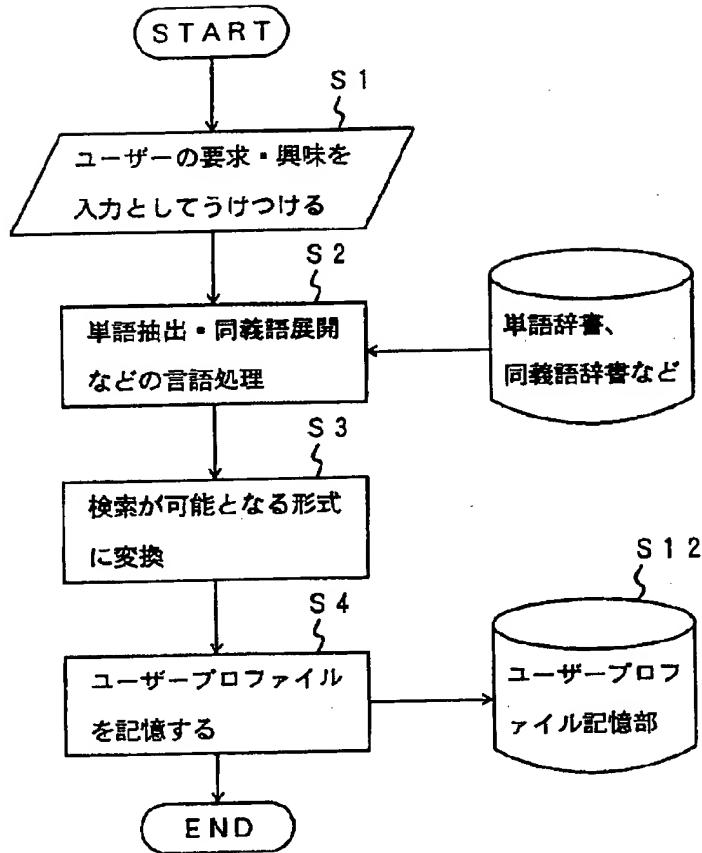
記事名	情報源
記事A	新聞社M
記事B	新聞社N
記事C	出版社O
記事D	新聞社M

ユーザー1に提示する記事

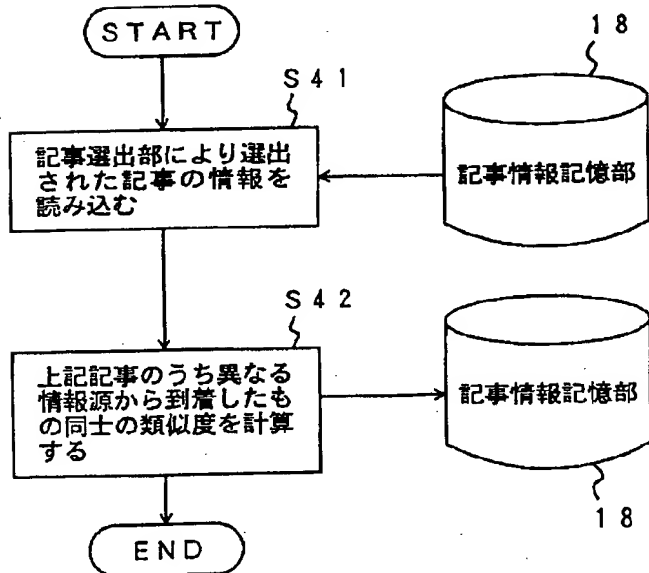
【図3】



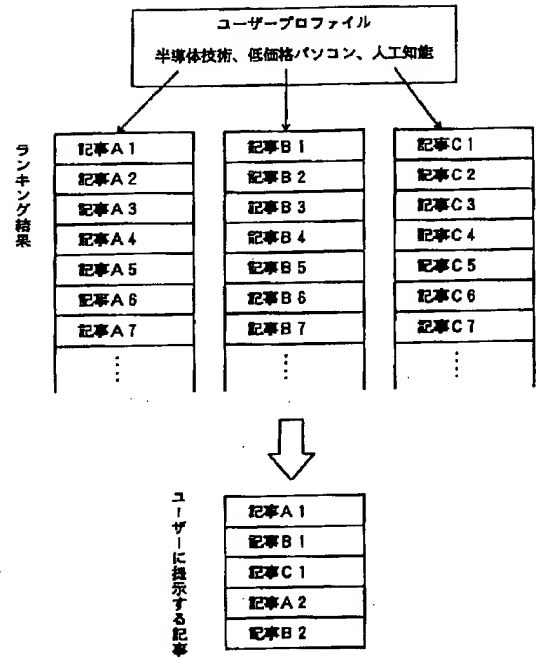
【図4】



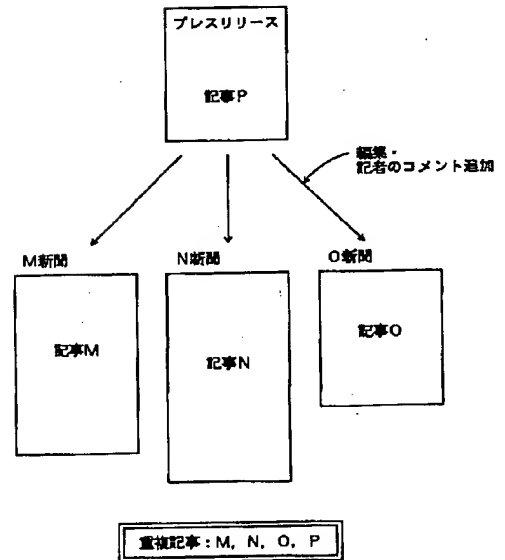
【図14】



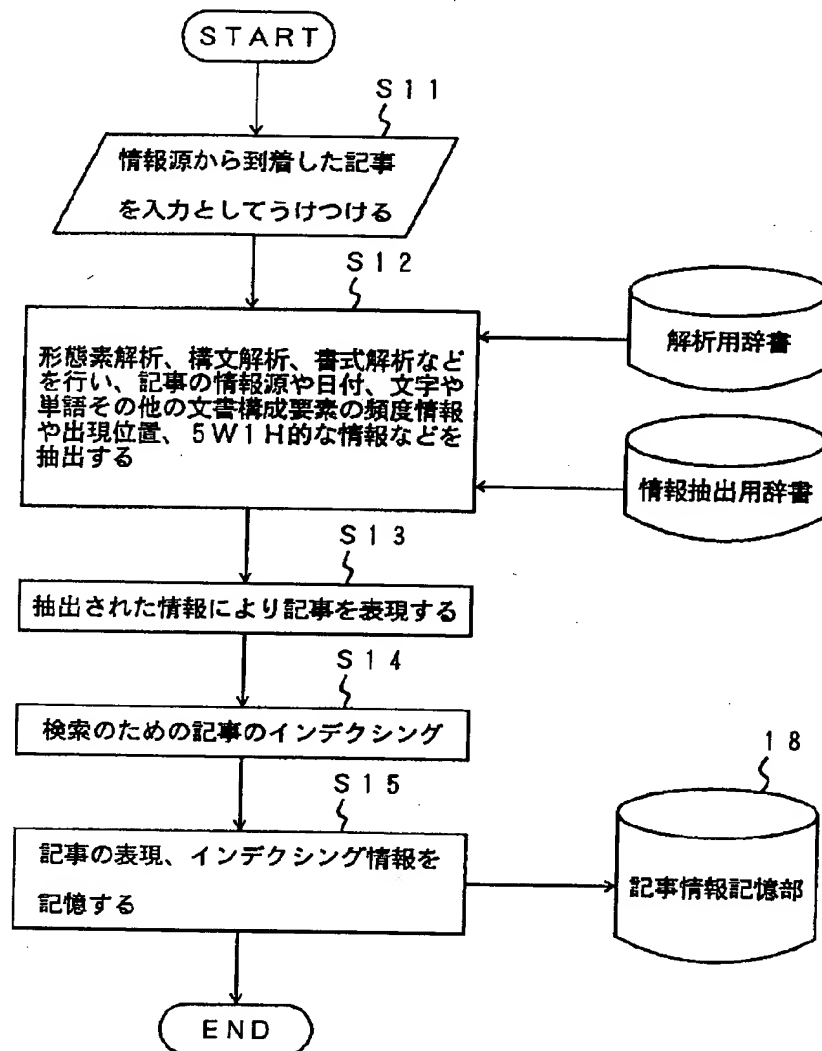
【図13】



【図17】

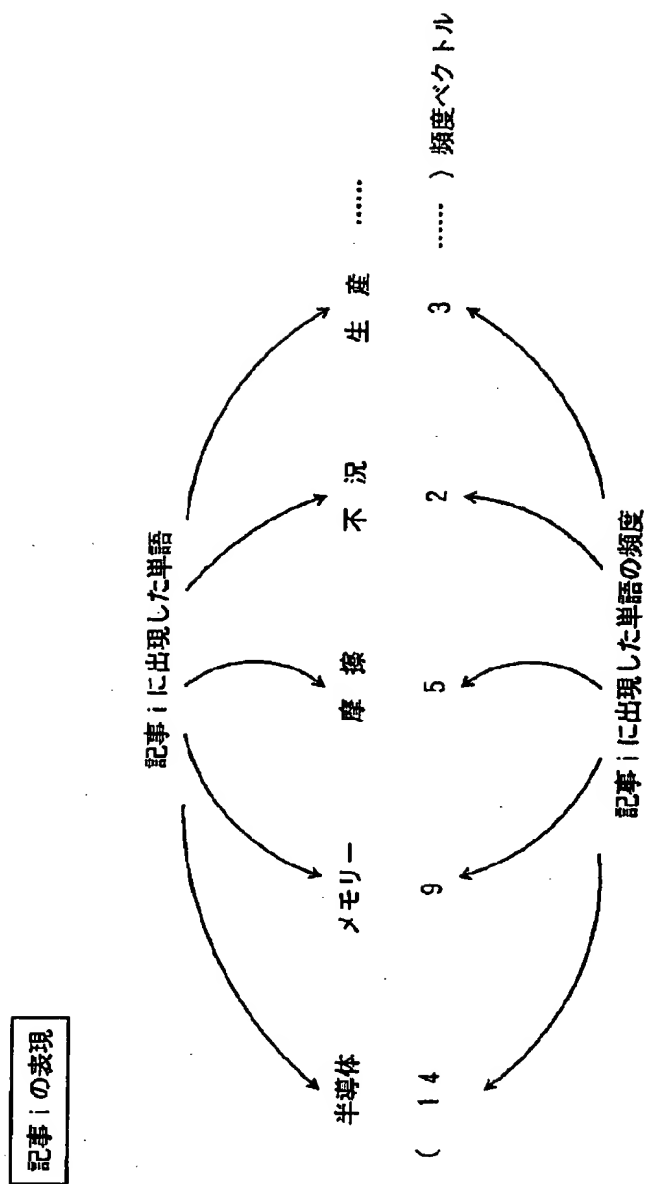


【図5】

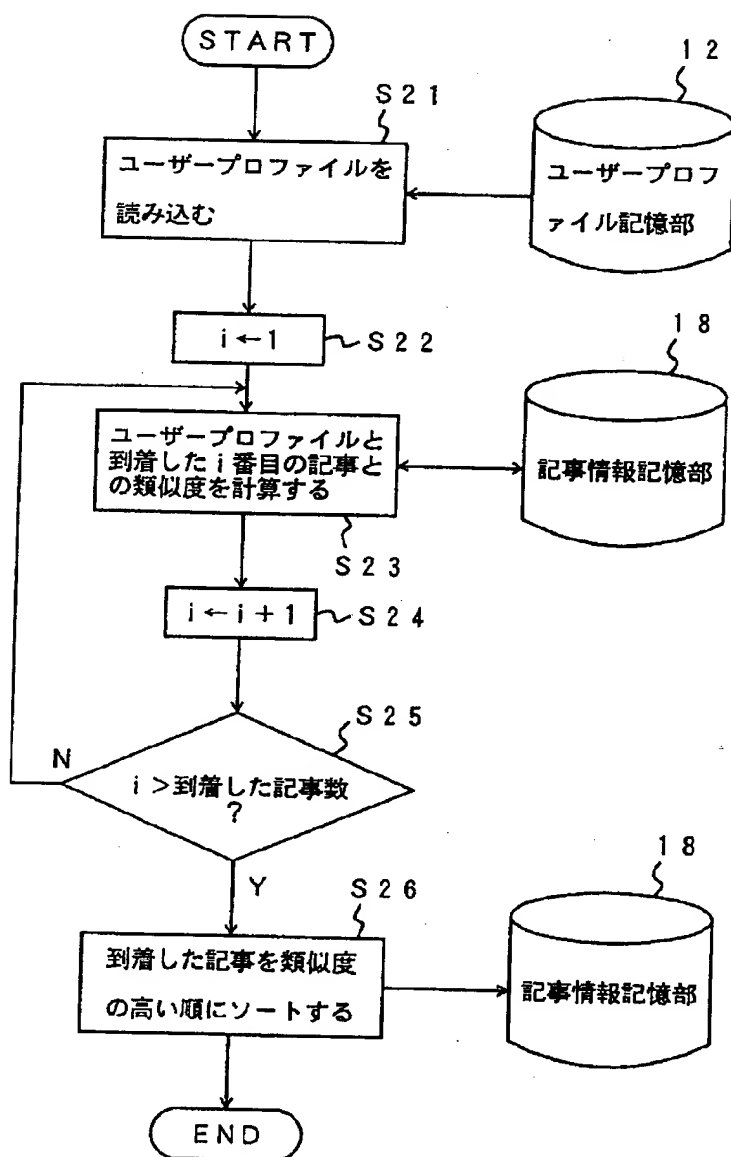


(31)

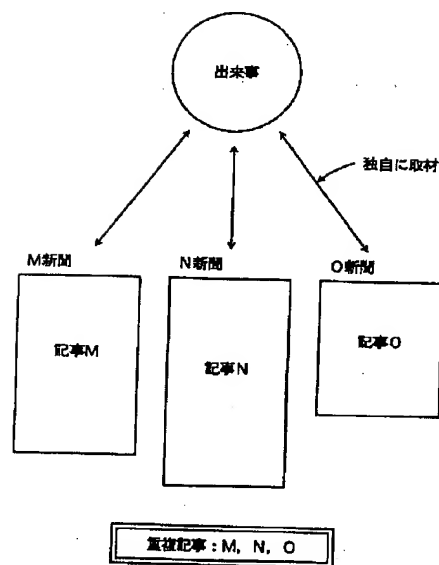
【図6】



【図8】



【図18】



【図19】

重複記事集合1

記事名	情報源	ユーザープロフィール と記事との類似度
記事A	新聞社M	0.71
記事C	新聞社O	0.84

(A)

重複記事集合2

記事名	情報源	ユーザープロフィール と記事との類似度
記事B	新聞社N	0.76
記事D	新聞社M	0.83

(B)

【図9】

順位	到着記事名	ユーザープロフィールとの類似度
1		0.97
2		0.96
3		0.96
4		0.94
5		0.91
6		0.89
7		0.89
8		0.87
9		0.85
10		0.79
11		0.79
12		0.64
13		0.63
14		0.62
15		0.61
16		0.60
17		0.60
18		0.59
19		0.57
20		0.40
...		

【図20】

関連記事（重複）：

元の記事との類似度

- 「情報サービスから撤退 ○×社」 0.98
△△新聞 朝刊 1面 ××文字
××年 ××月 15日
記事ID: △△△△△
- 「○× 情報サービスを断念」 0.92
□□新聞 朝刊 ××文字
××年 ××月 15日
記事ID: △△△△△

記事見出し: 「○×社 情報サービスビジネスから撤退」

情報源: ○○新聞 朝刊 1面 ××文字
××年 ××月 15日

記事ID: △△△△△

本文:

付加情報

本文情報

【図22】

元の記事／関連記事：

- 半導体協会の..... (元の記事)
○○新聞 朝刊 経済面1
××年 ××月 15日
記事ID: △△△△△
- <関連記事2の記事見出し>
△○新聞 朝刊 1面
××年 ××月 ××日
記事ID: △△△△△

記事見出し: <関連記事1の記事見出し>

情報源: 週刊○○ ページ××
××年 ××月 ××日

記事ID: △△△△△

本文:

付加情報

本文情報

【図21】

関連記事：

- <関連記事1の記事見出し>
週刊○○ ページ××
××年 ××月 ××日
記事ID: △△△△△
- <関連記事2の記事見出し>
△○新聞 朝刊 1面
××年 ××月 ××日
記事ID: △△△△△

記事見出し: 「半導体協会の.....」

情報源: ○○新聞 朝刊 経済面1
××年 ××月 15日

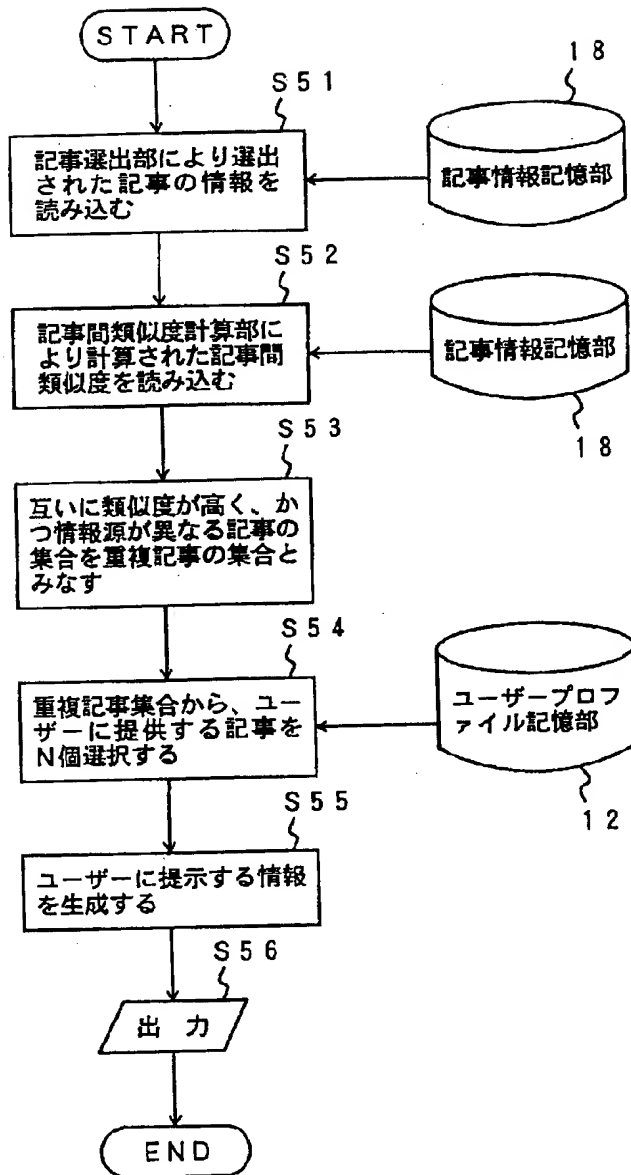
記事ID: △△△△△

本文:

付加情報

本文情報

【図16】



【図72】

ユーザーが選択しているトピック	重要度
トピックA	2.0
トピックB	1.5
トピックC	1.0
トピックD	0.5

【図26】

ユーザープロフィールとの類似度	
1. 記事1の見出し	0.97
○○新聞 朝刊 2面 ××文字	
2. 記事2の見出し	0.84
△△新聞 1面 ××文字	
3. 「○×社 情報サービスビジネスから撤退」	0.79
○○新聞 朝刊 1面 ××文字	
《重複記事》	
3.1 「情報サービスから撤退 ○×社」	0.77
△△新聞 朝刊 1面 ××文字	
《元の記事との類似度》	
3.2 「○× 情報サービスを断念」	0.77
□□新聞 朝刊 ××文字	
《元の記事との類似度》	
4. 記事4の見出し	0.78
週刊△○ ××ページ	

【図28】

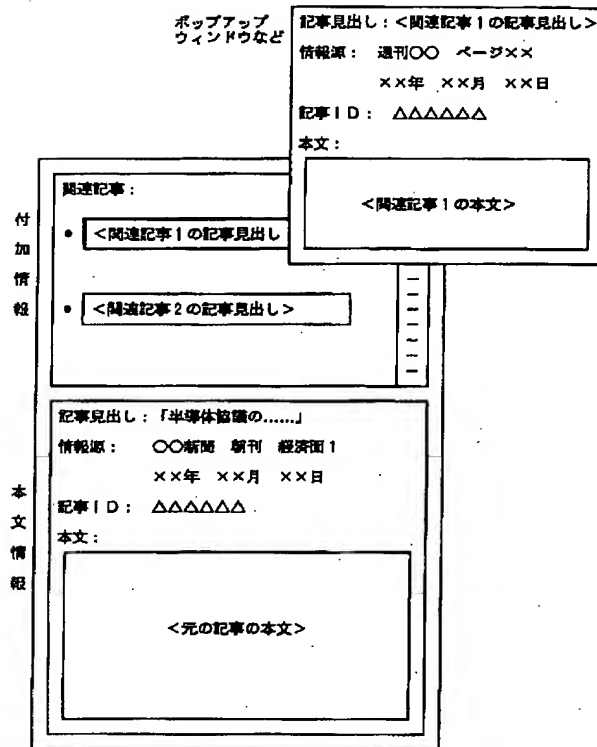
前回ユーザーに提示された記事	記事名	情報源	ユーザープロフィールとの類似度
	記事A	新聞社M	0.81
	記事B	新聞社M	0.62
	記事C	新聞社N	0.79
	記事D	出版社O	0.72

(A)

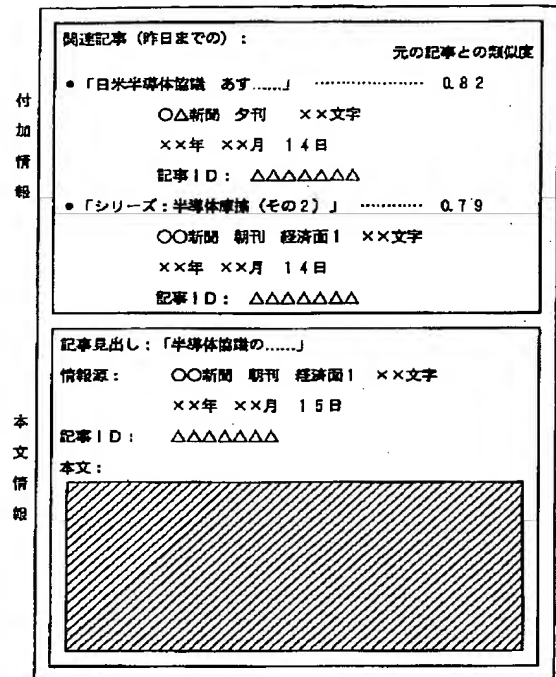
今回記事選出部により選出された記事	記事名	情報源	ユーザープロフィールとの類似度
	記事E	新聞社M	0.84
	記事F	新聞社N	0.72
	記事G	新聞社N	0.81
	記事H	出版社O	0.77

(B)

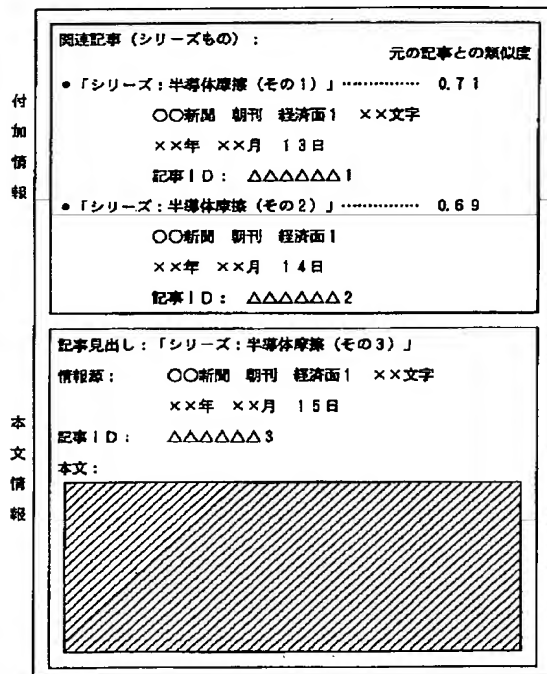
【図23】



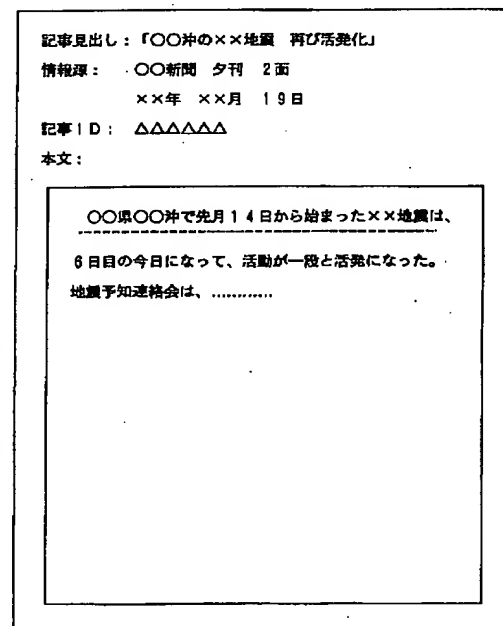
【図30】



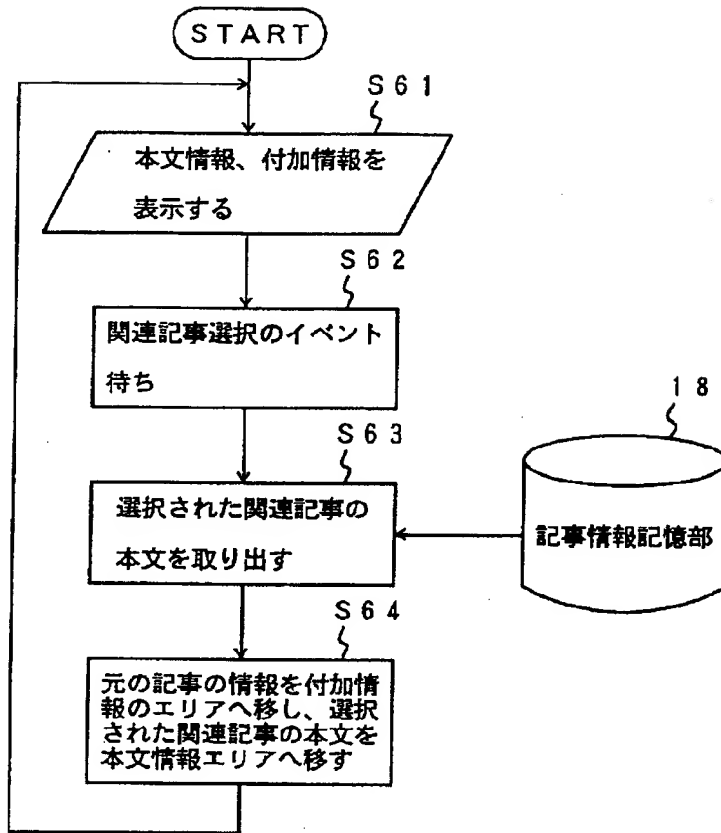
【図31】



【図32】



【図24】



【図39】

検索結果			ユーザーに提示する記事及びその順序		
順位	記事名	ユーザープロフィールとの類似度	提示順序	記事名	
1	記事1	0.98	1	記事1	半導体技術
2	記事2	0.97	2	記事5	
3	記事3	0.89	3	記事7	
4	記事4	0.89	4	記事2	低価格パソコン
5	記事5	0.85	5	記事3	
6	記事6	0.84	6	記事6	人工知能
7	記事7	0.80	7	記事4	
8	記事8	0.80	8	記事8	
9	記事9	0.79			
10	記事10	0.61			
⋮	⋮	⋮			

(a)

提示順序	記事名	
1	記事1	半導体技術
2	記事5	
3	記事7	
4	記事2	低価格パソコン
5	記事3	
6	記事6	人工知能
7	記事4	
8	記事8	

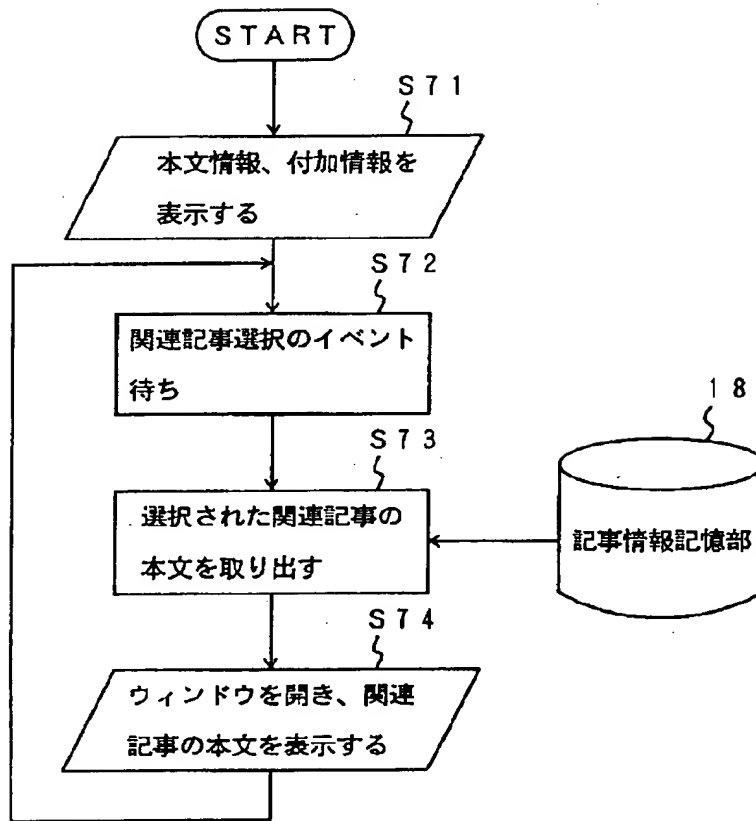
(b)

【図43】

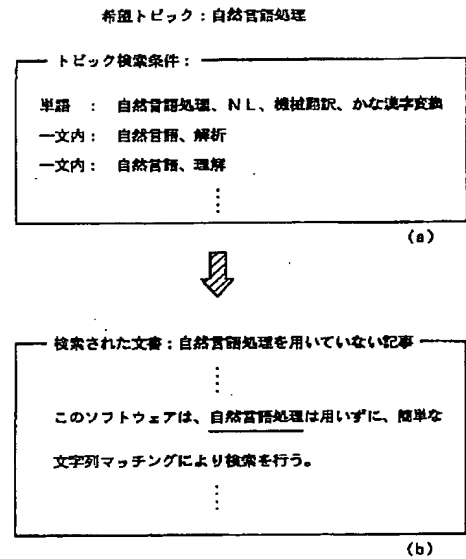
トピック1に対する検索結果

順位	到着記事名	トピックとの類似度
1		0.97
2		0.96
3		0.96
4		0.94
5		0.91
6		0.89
7		0.89
8		0.87
9		0.85
10		0.79
11		0.64
⋮	⋮	⋮

【図25】



【図51】



【図45】

ユーザーの選択したトピック

トピック1 半導体貿易	トピック2 低価格パソコン	トピック3 人工知能
1 記事1-1	1 記事2-1	1 記事3-1
2 記事1-2	2 記事2-2	2 記事3-2
3 記事1-3	3 記事2-3	3 記事3-3
4 記事1-4	4 記事2-4	4 記事3-4
5 記事1-5	5 記事2-5	5 記事3-5
6 記事1-6	6 記事2-6	6 記事3-6
7 記事1-7	7 記事2-7	7 記事3-7
8 記事1-8	8 記事2-8	8 記事3-8
9 記事1-9	9 記事2-9	9 記事3-9
10 ...	10 ...	10 ...

検索結果

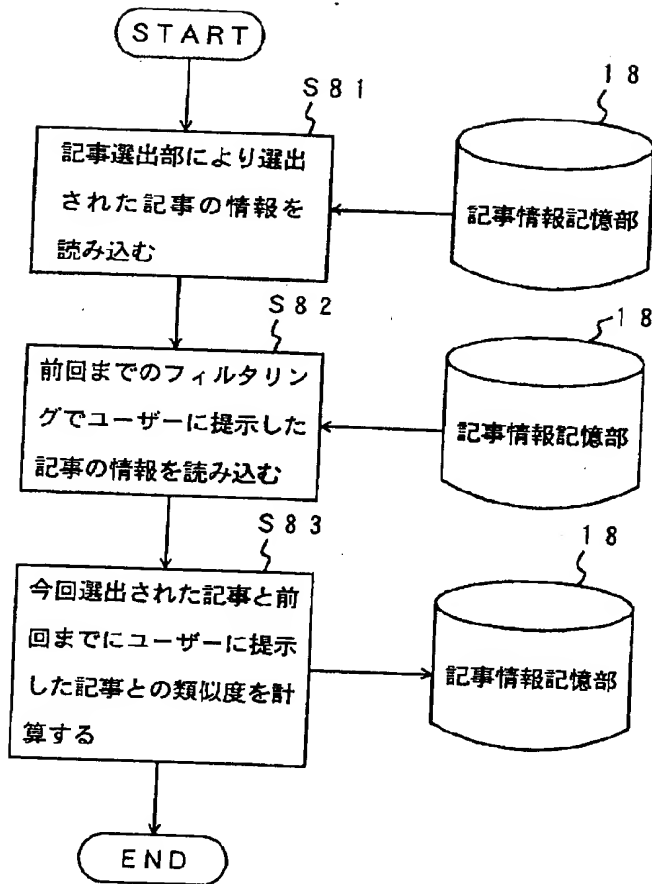
【図47】

今回到着した記事（記事見出しのリスト）：

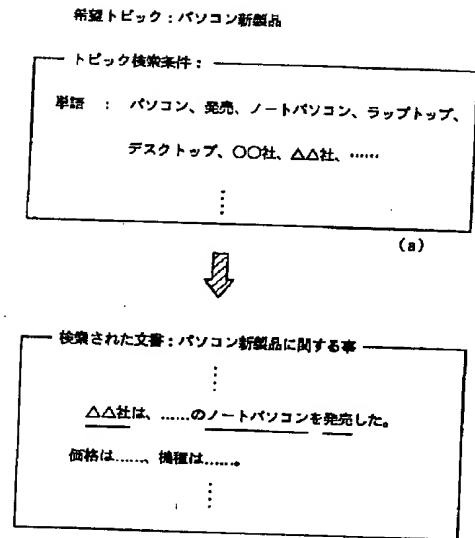
記事番号	記事見出し	情報源	該当トピック	トピックと記事との類似度
1.	記事見出し1	〇〇新聞	半導体貿易	1.92
2.	記事見出し2	〇〇新聞	半導体貿易	1.30
3.	記事見出し3	△△新聞	半導体貿易	1.01
4.	記事見出し4	〇〇新聞	低価格パソコン	1.20
5.	記事見出し5	週刊××	低価格パソコン	0.91
6.	記事見出し6	□□新聞	半導体貿易/低価格パソコン	1.05/0.80

計6件

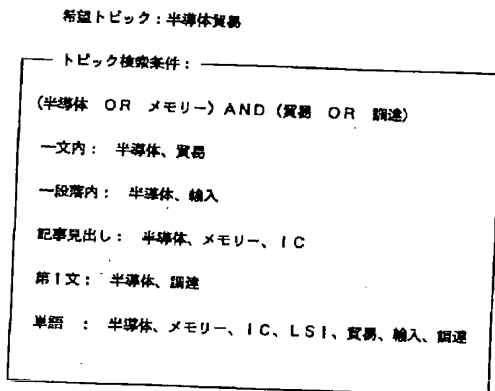
【図27】



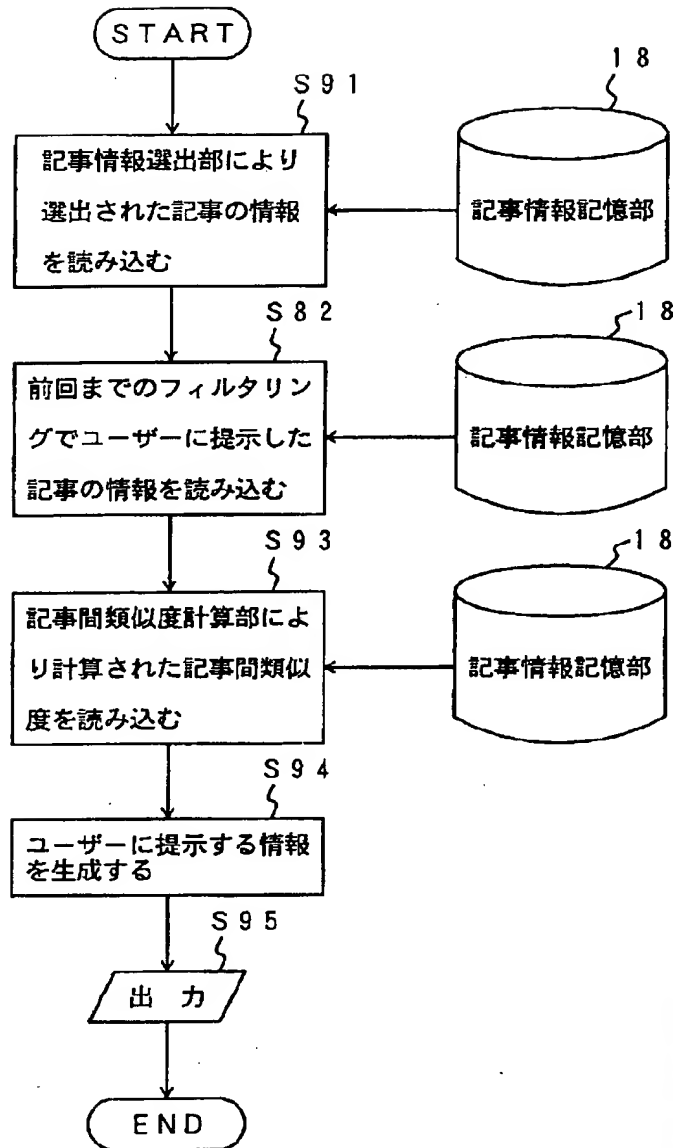
【図53】



【図54】



【図29】



【図59】

	ユーザー1	ユーザー2	ユーザー3	ユーザー4
記事1	送信する		送信する	送信する	
記事2	送信する	送信する		送信する	
記事3		送信する			
記事4		送信する			
記事5				送信する	
.....					

【図34】

記事見出し: 「〇〇沖で地震 マグニチュード4」

情報源: 〇〇新聞 夕刊 1面
××年 ××月 14日

記事ID: △△△△△

本文:

〇〇県の〇〇沖で今日明け方から地震が
相次いだ。 〇〇沖の地震は昨年××月以来
××ヶ月ぶり、気象庁は.....

【図37】

付加情報

本文情報

関連記事 (今日の):

元の記事との類似度

- 「××社 半導体シェア抜く.....」 0.80
週刊〇〇 ページ×× ××文字
××年 ××月 15日
記事ID: △△△△△
- 「××半導体工場.....生産開始」 0.70
△〇新聞 朝刊 1面 ××文字
××年 ××月 15日
記事ID: △△△△△

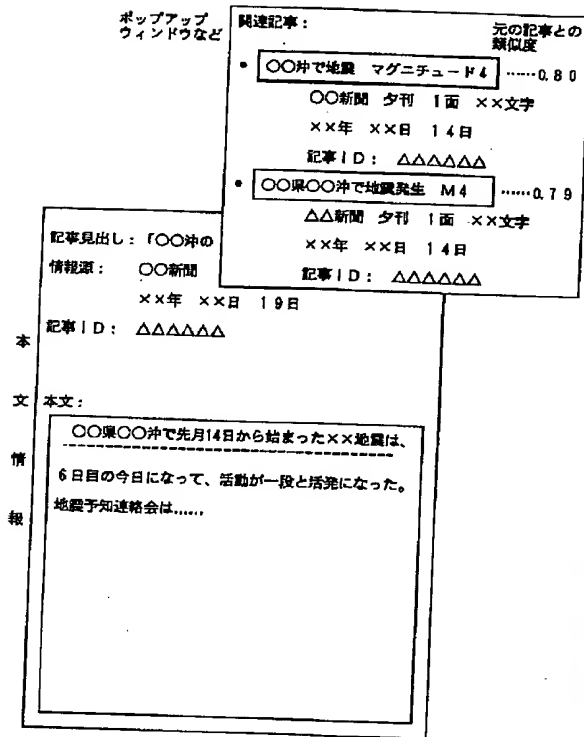
記事見出し: 「半導体協会の.....」

情報源: 〇〇新聞 朝刊 経済面1 ××文字
××年 ××月 15日

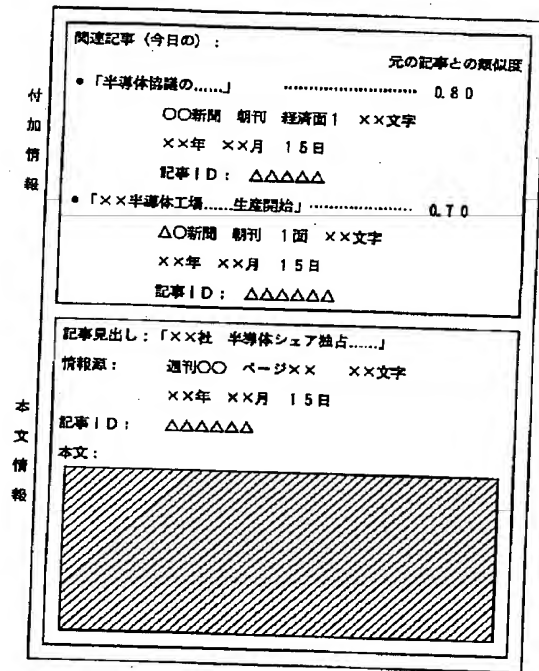
記事ID: △△△△△

本文:

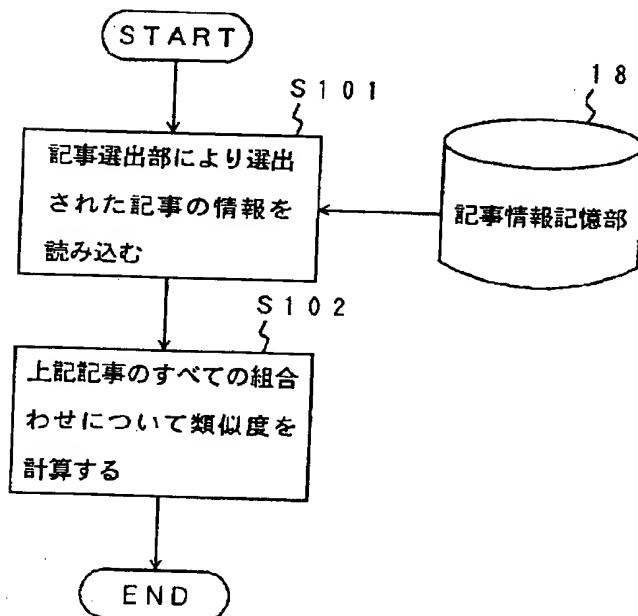
【図33】



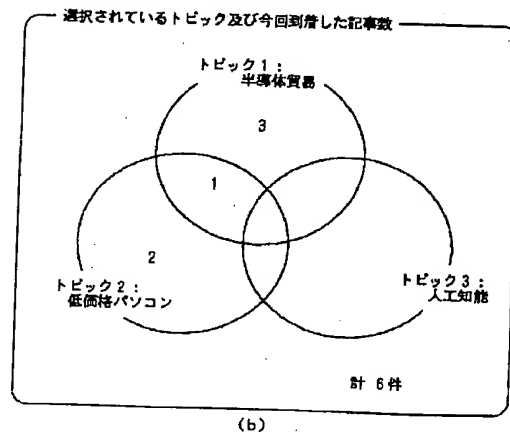
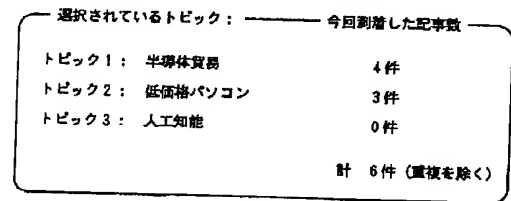
【図38】



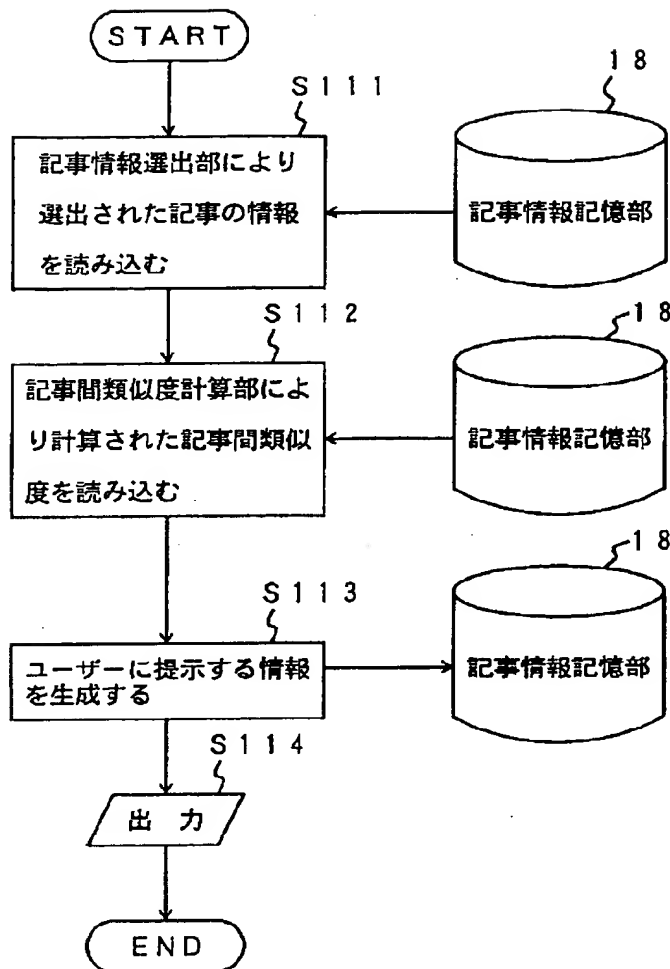
【図35】



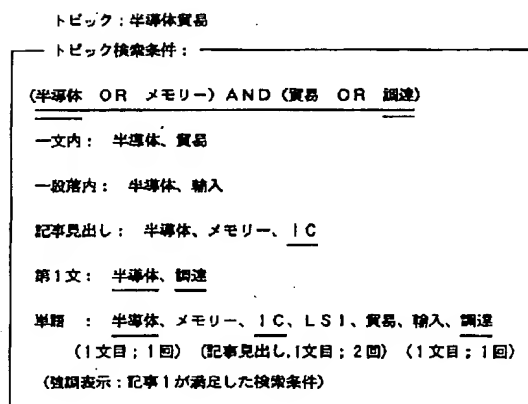
【図48】



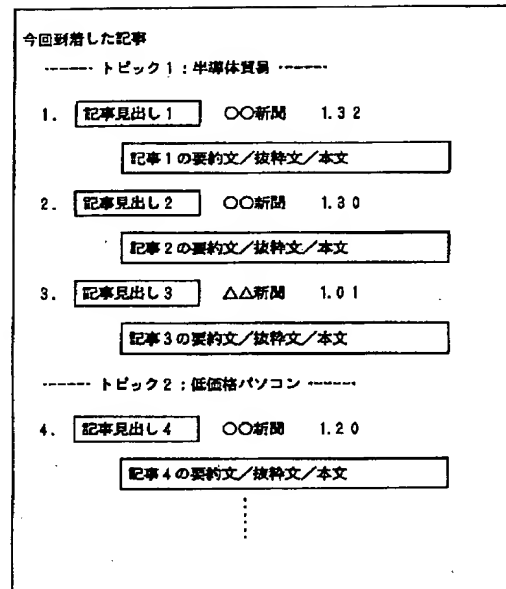
【図36】



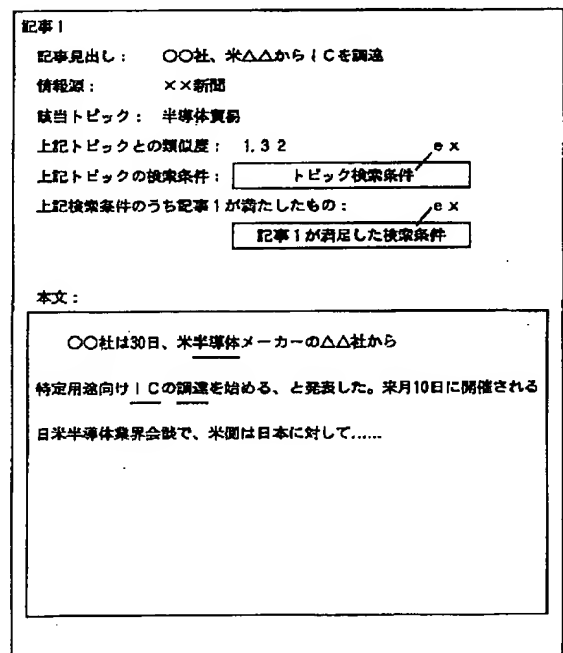
【図56】



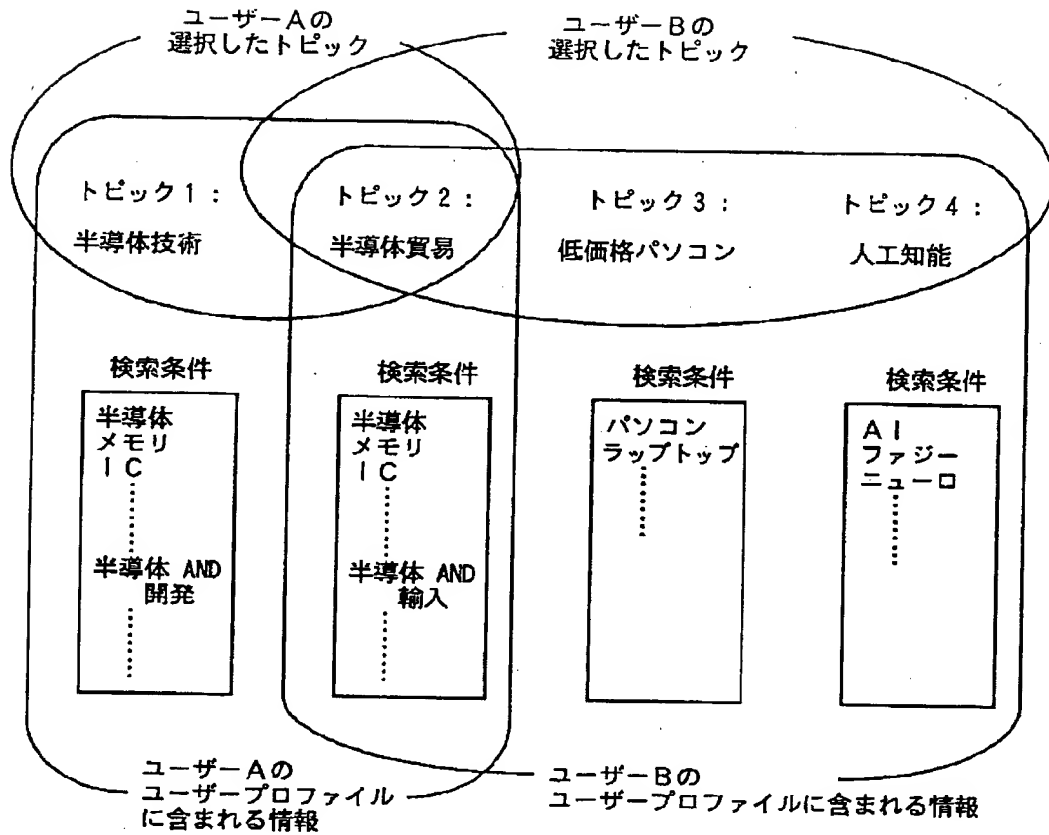
【図49】



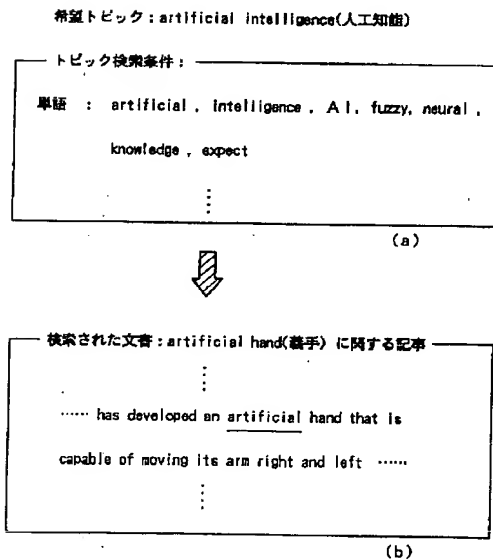
【図50】



【図40】



【図52】

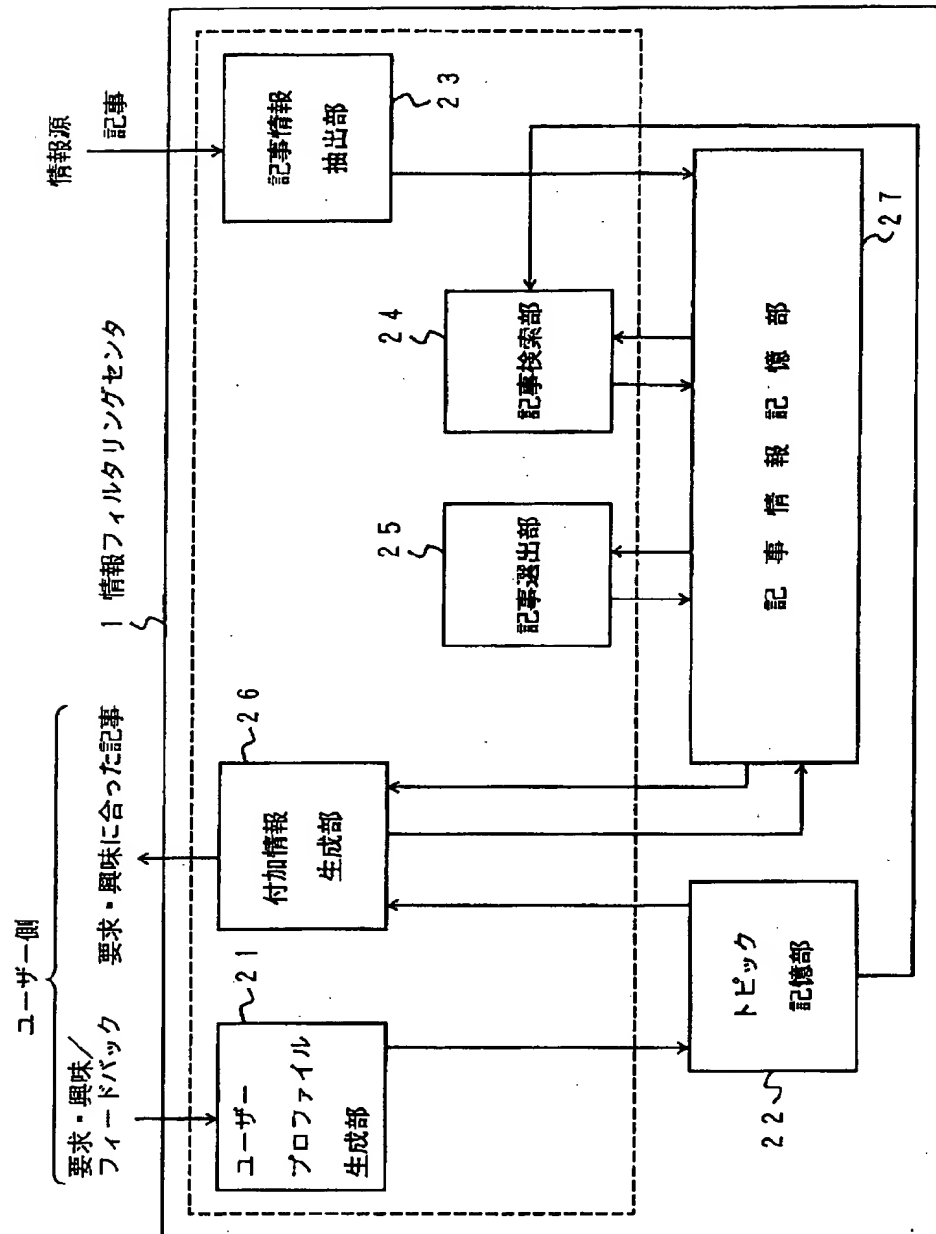


【図60】

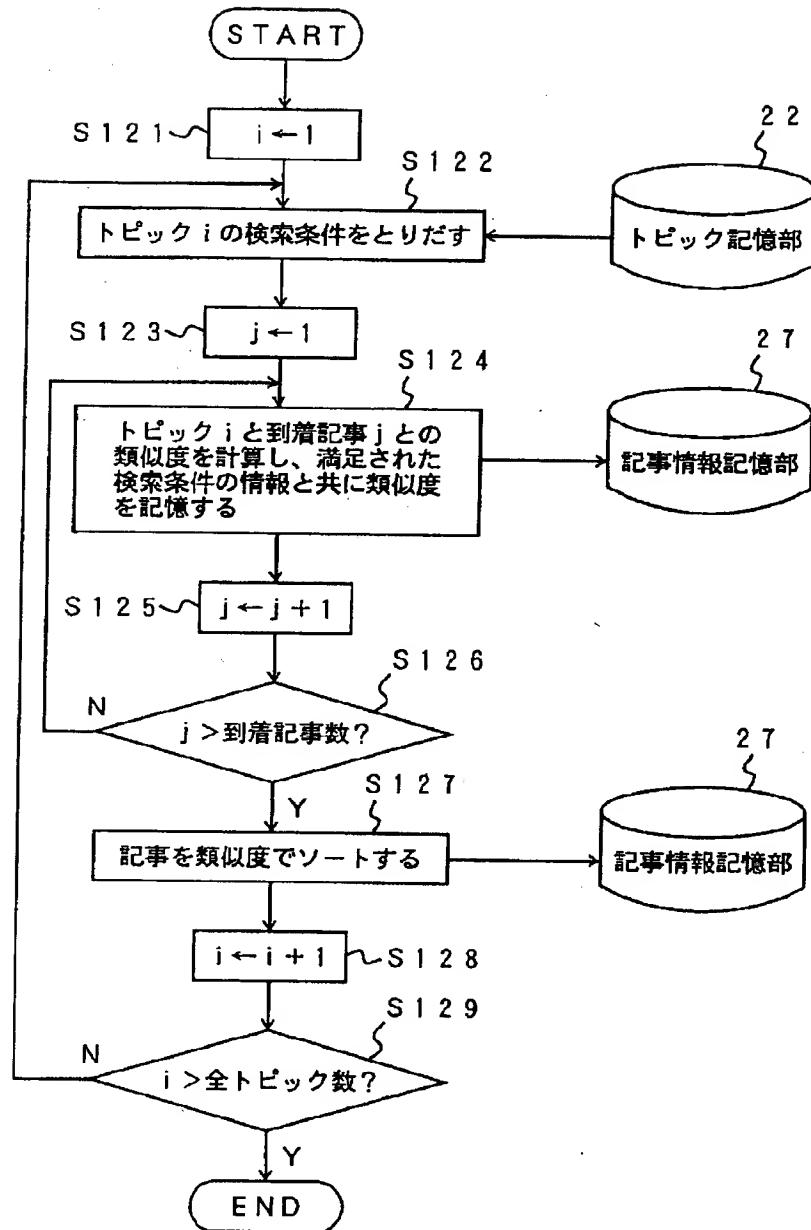
今回到着した記事(記事見出しのリスト):

記事番号	記事見出し	情報源	受信ユーザー数/全ユーザー数
1.	記事見出し1	〇〇新聞	250人/4,000人
2.	記事見出し2	〇〇新聞	250人/4,000人
3.	記事見出し3	△△新聞	200人/4,000人
4.	記事見出し4	〇〇新聞	1,000人/4,000人
5.	記事見出し5	週刊××	500人/4,000人
6.	記事見出し6	□□新聞	300人/4,000人

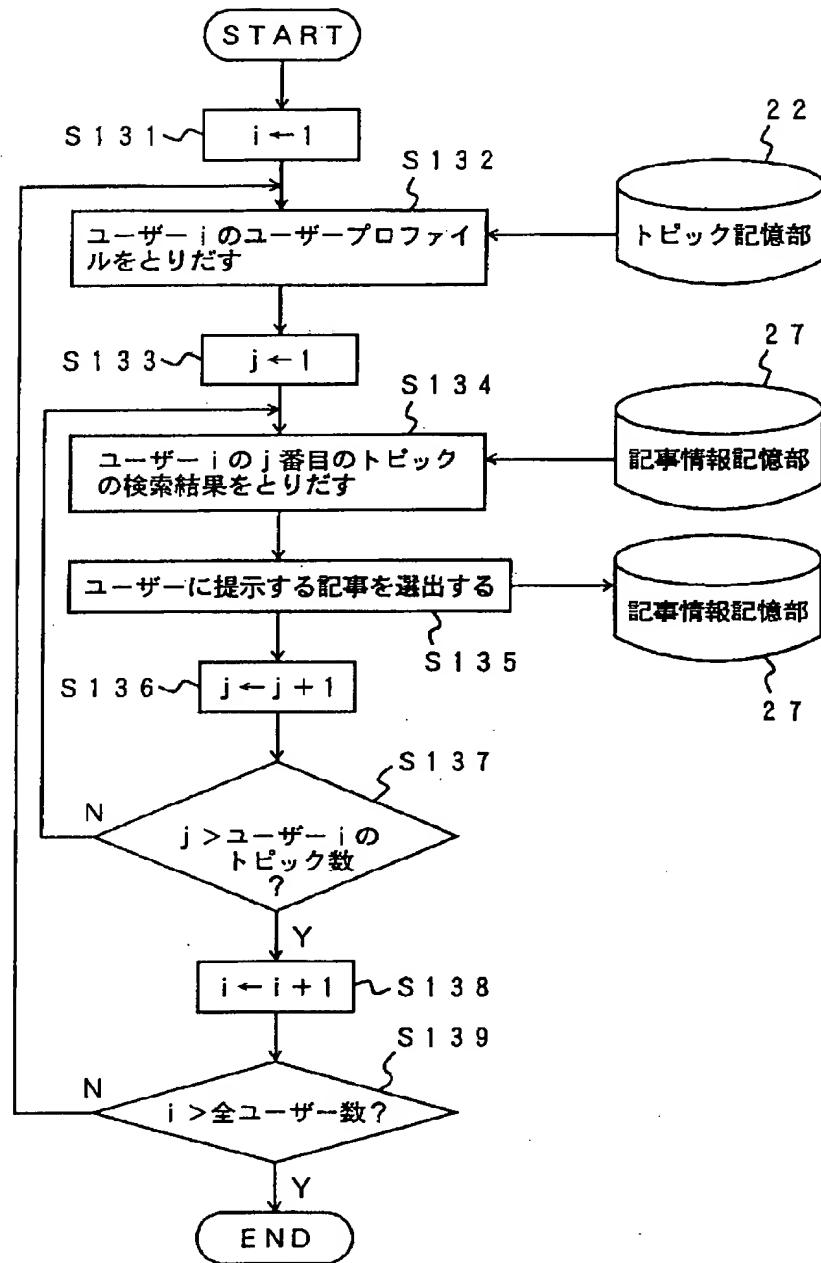
【図41】



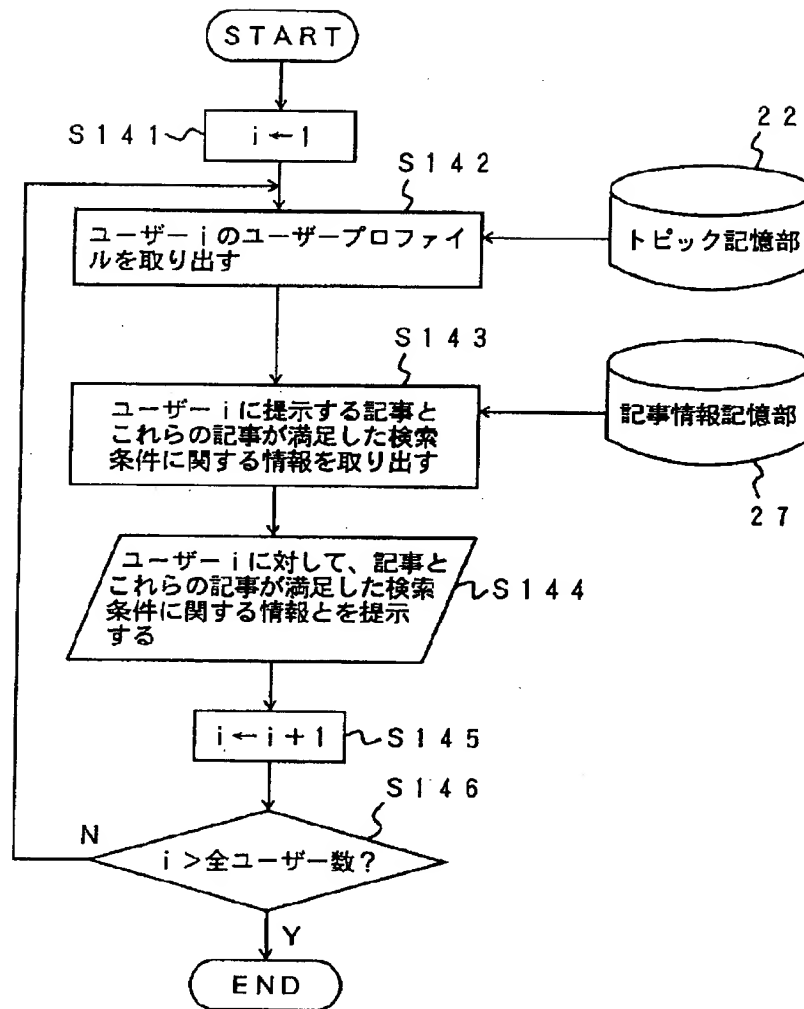
【図42】



【図44】



【図46】



【図70】

ユーザー i に提示する記事	適合したトピック
記事1	トピックA
記事2	トピックA
記事3	トピックB
記事4	トピックC、トピックD

【図75】

ユーザー i に提示する記事	属性 (日付)
記事1	××年5月26日
記事2	××年5月23日
記事3	××年5月23日
記事4	××年5月20日

【图 5 5】

該当トピック：半導体貿易

上記トピックとの類似度：1.23

トピック検索条件:

(半導体 OR メモリー) AND (貿易 OR 調達)

一文內： 半導體、貿易

一段落內： 半導體、輸入

記事見出し： 半導体、メモリー、IC

第1文： 半導体、関連

単語 : 半導体、メモリー、IC、LSI、貿易、輸入、調達

記事1が満足した検索条件

〈半導体 OR メモリー〉 AND 〈貿易 OR 製造〉

記事見出し： IC (1回)

第1文： 半礫体（1回）、調速（1回）

単語： 半導体、IC、隣邊
(1文目；1回) (記事見出し、(1文目；1回)
1文目；
2回)

【図 6 1】

今回到着した記事（要約文／抜粋文）

1. 記事見出し1 ○○新聞 250人/4,000人

記事1の要約文/抜粋文

2. 記事見出し 2 ○○新聞 250人／4,000人

記事2の要約文／抜粋文

3. 記事見出し 3 △△新聞 200人／4,000人

記事3の要約文／抜粋文

4. 記事見出し 4 ○○新聞 1,000人/4,000人

記事4の要約文/抜粋文

.....

【図 6 3】

【图 6 2】

記事 1

記事見出し： ○○社、米△△社から C を買収

情報源： ××新聞

該当トピック： 半導体貿易

受領ユーザー数/全ユーザー数： 250人/4,000人

本文：

本文情報

・記事1

記事見出し： ○○社、米△△社からI Cを調達

情報源： ××新聞

該当トピック： 半導体貿易

受講ユーザー数/全ユーザー数： 250人/4,000人

《内記》

性別：	男	性	150人
	女	性	100人

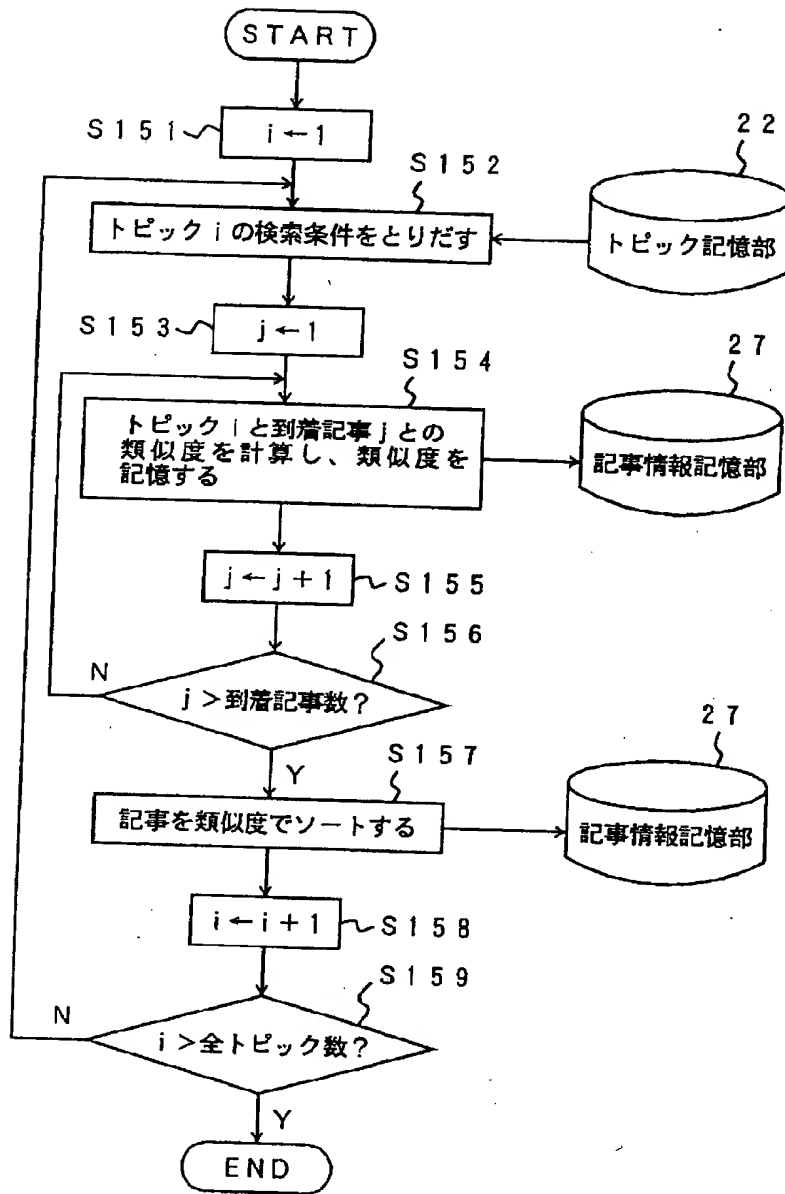
国別： 日本 200人
米 30人
その他 20人

トピック別： 半導体貿易 180人
IC 50人
両方 20人

本文：

本文情報

【図57】



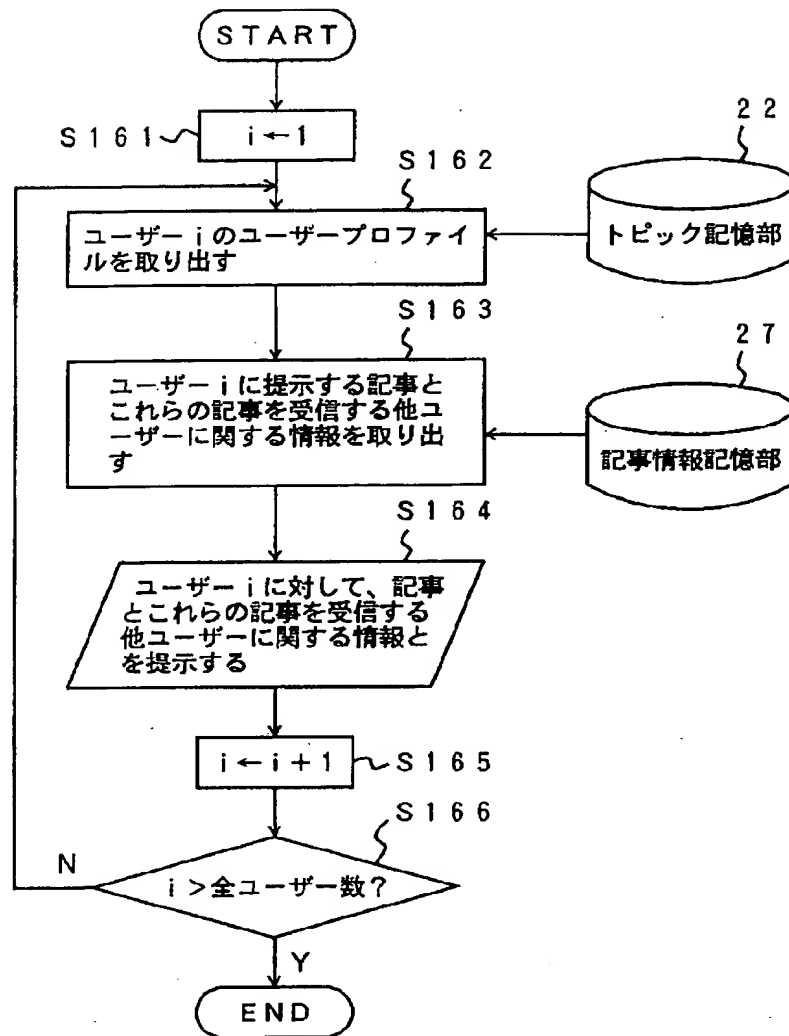
【図84】

記事1タイトル	8/4
□ 記事2タイトル	8/4
記事3タイトル	8/4
記事4タイトル	8/4
記事5タイトル	8/4
△ 記事6タイトル	8/4
記事7タイトル	8/3

【図77】

ユーザー i に提示する記事	属性 (新聞社)
記事1	○○新聞
記事2	△△新聞
記事3	△△新聞
記事4	××新聞

【図58】



【図86】

1995/08/04	
01:	マルチメディアの市場規模1兆622.5億円、昨年、前年比3.5%増——振興協 日本○○新聞 p.11 513字
02:	昨年、ソフト振興協まとめ、マルチメディア市場3.5%の伸び ××産業新聞 p.6 452字
03:	マルチメディア事業が3.5%増昨年 ○○新聞 p.12 351字

【図64】

前回のrelevance feedback情報:

記事番号	記事見出し	情報源	前回の判定	有用	やや有用	不要	計
a 1.	記事見出し1	〇〇新聞	有用	100人	100人	50人	250人
a 2.	記事見出し2	〇〇新聞	有用	200人	50人	0人	250人
a 3.	記事見出し3	△△新聞	不要	0人	150人	50人	200人
a 4.	記事見出し4	〇〇新聞	やや有用	800人	300人	100人	1,000人
a 5.	記事見出し5	週刊××	不要	100人	0人	400人	500人
a 6.	記事見出し6	□□新聞	有用	200人	100人	0人	300人

今回到着した記事:

b 1.

b 2.

b 3.

b 4.

【図65】

前回のrelevance feedback情報:

記事番号	記事見出し	情報源	前回の判定	前回の他ユーザーの判定 平均
a 1.	記事見出し1	〇〇新聞	10点	8.9点 (250人)
a 2.	記事見出し2	〇〇新聞	9点	9.0点 (250人)
a 3.	記事見出し3	△△新聞	0点	4.0点 (200人)
a 4.	記事見出し4	〇〇新聞	4点	5.2点 (1,000人)
a 5.	記事見出し5	週刊××	1点	7.4点 (500人)
a 6.	記事見出し6	□□新聞	9点	9.9点 (300人)

今回到着した記事:

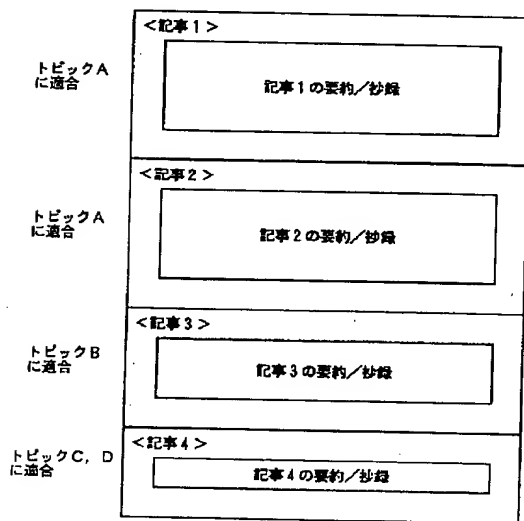
b 1.

b 2.

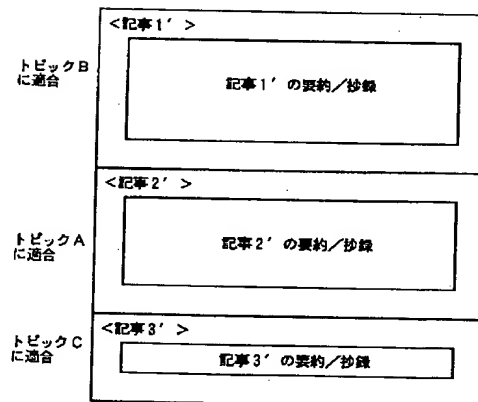
b 3.

b 4.

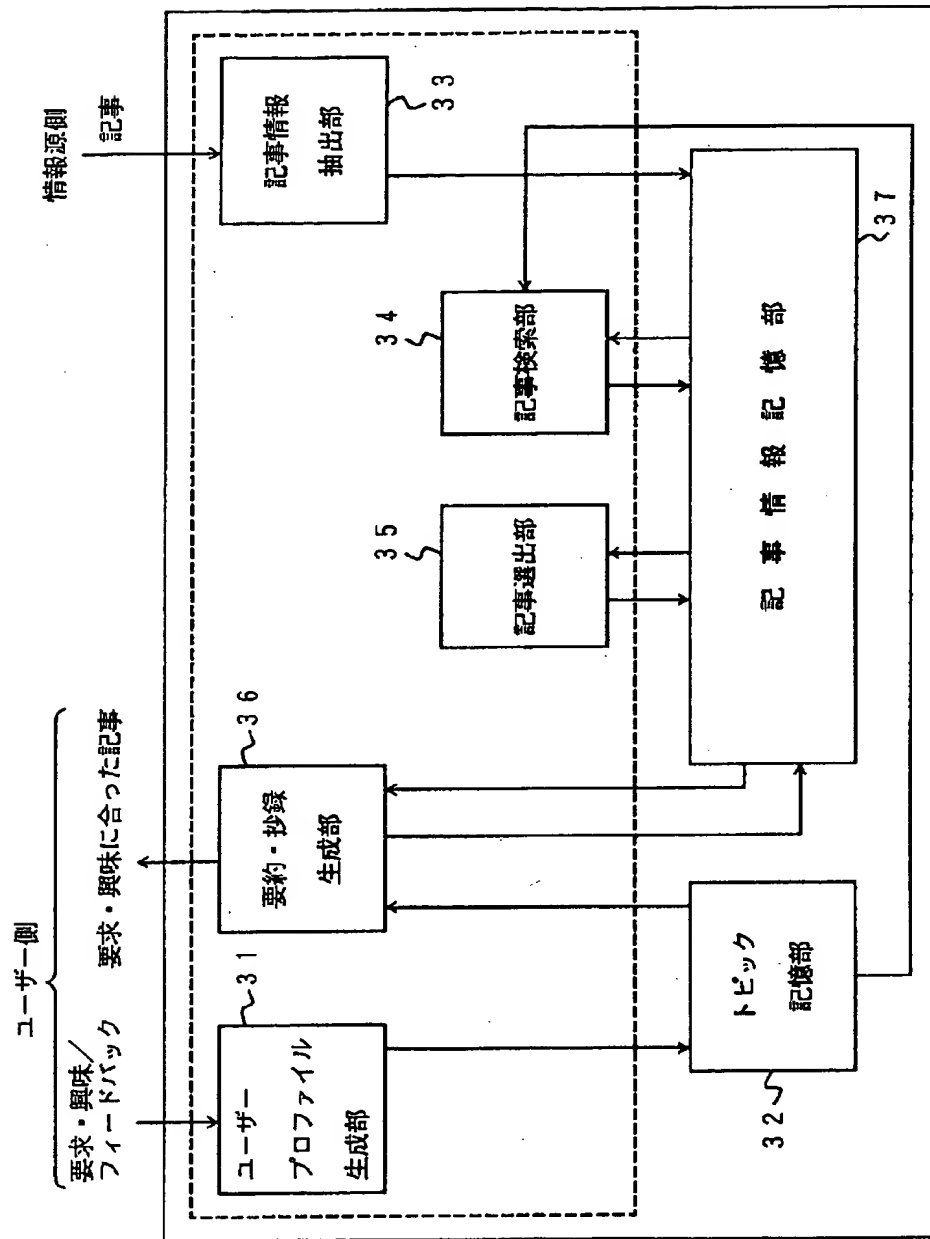
【図71】



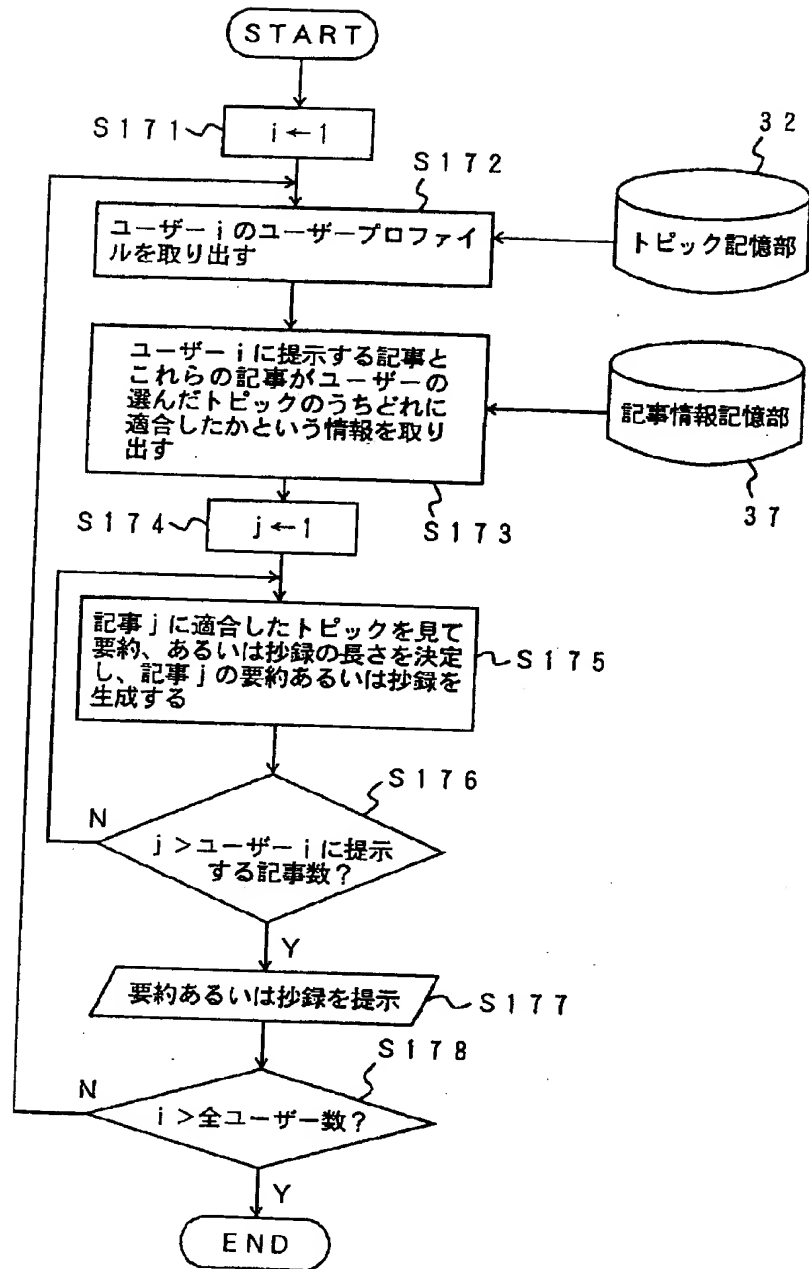
【図73】



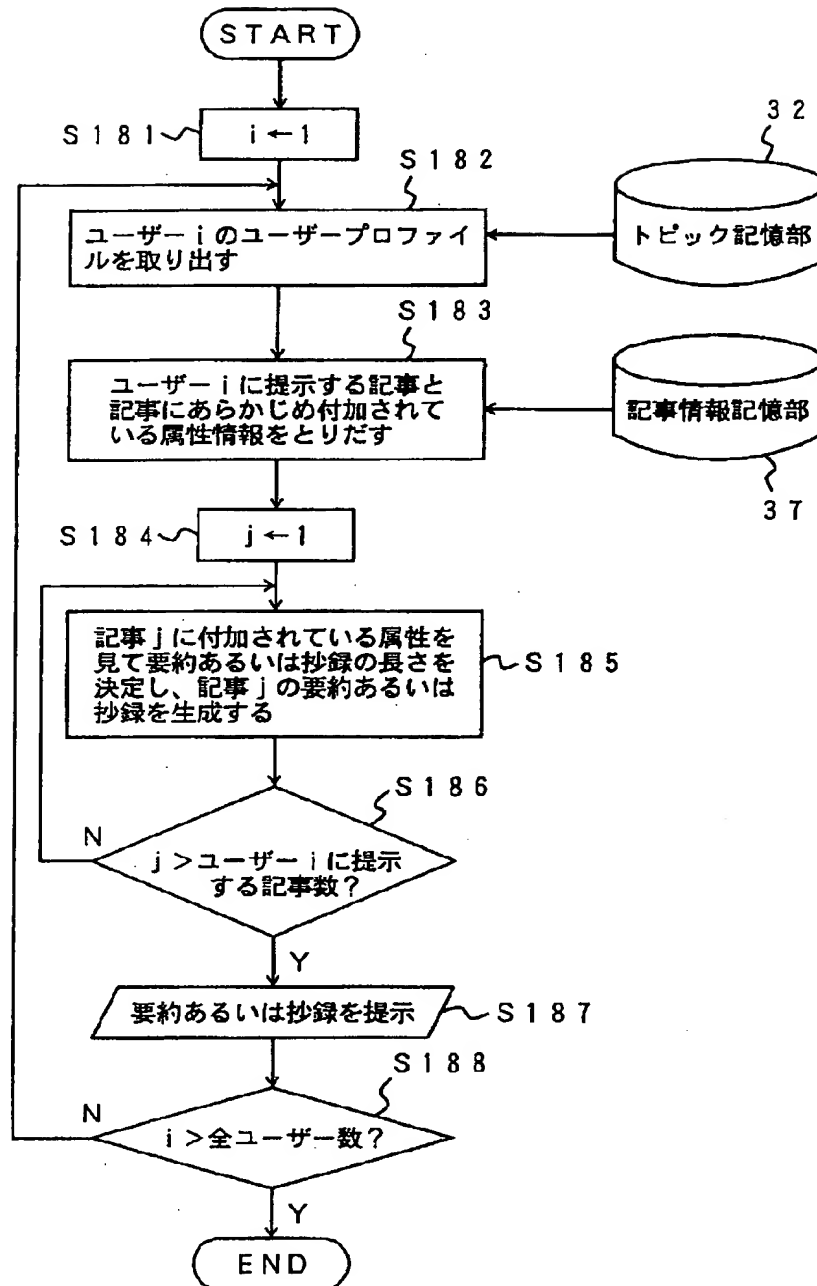
【図66】



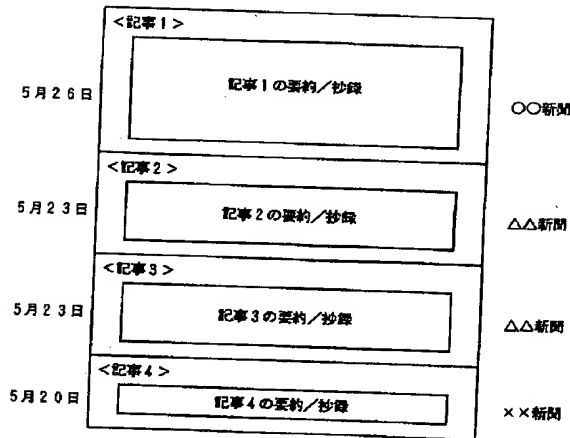
【図68】



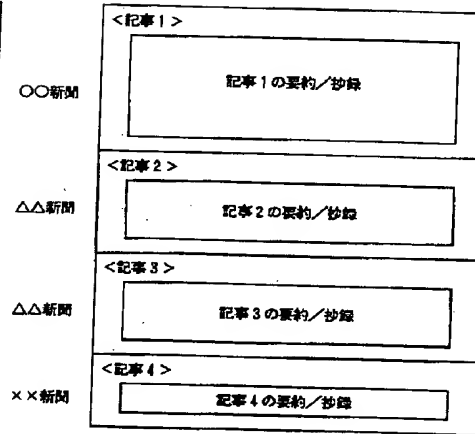
【図74】



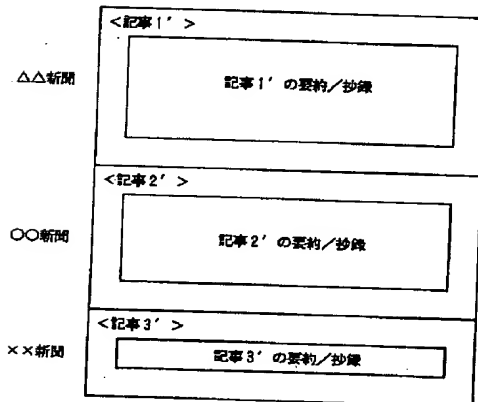
【図76】



【図78】



【図79】



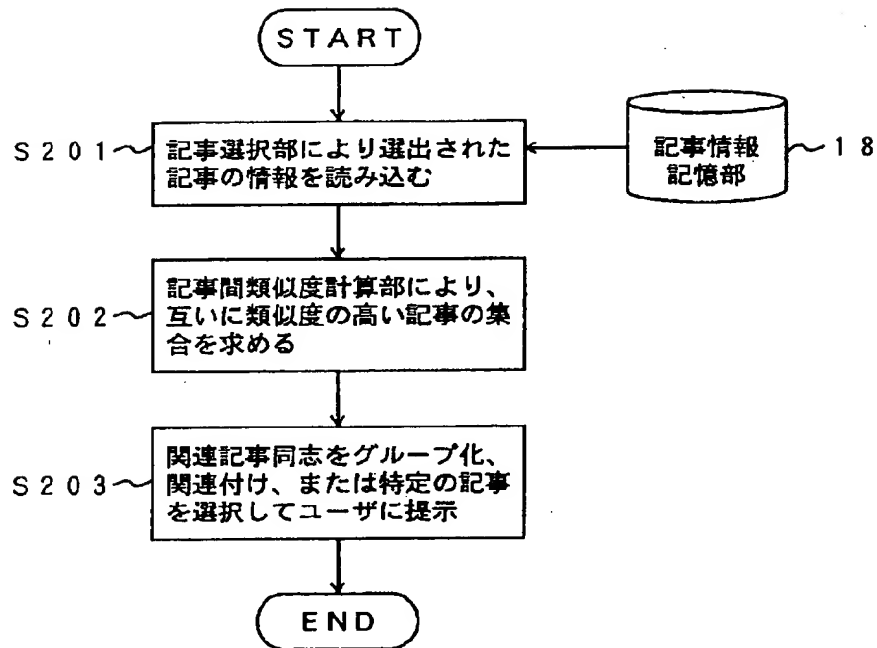
【図85】

001:	マルチメディアの市場規模1兆6225億円、昨年、前年比35%増——振興協
001:	日本○○新聞 朝刊 p.11 513字
002:	第4部マルチメディアを取り込め(1)航空、情報網で大連合(サービス産業)
	××産業新聞 p.15 1884字
003:	<魚眼図>コンテンツ産業
	○○新聞夕刊 p.6 704字
004:	独のマルチメディア実験/日本人が初参加
	××新聞 朝刊 p.13 388字
005:	☆マルチメディア市場急成長、94年35%増1兆6千億円
	×××新聞朝刊 p.3 0字
006:	マルチメディアの課題電子情報通信学会報告から(8)情報の貴重格差(註)
△02	○○産業新聞 p.5 844字

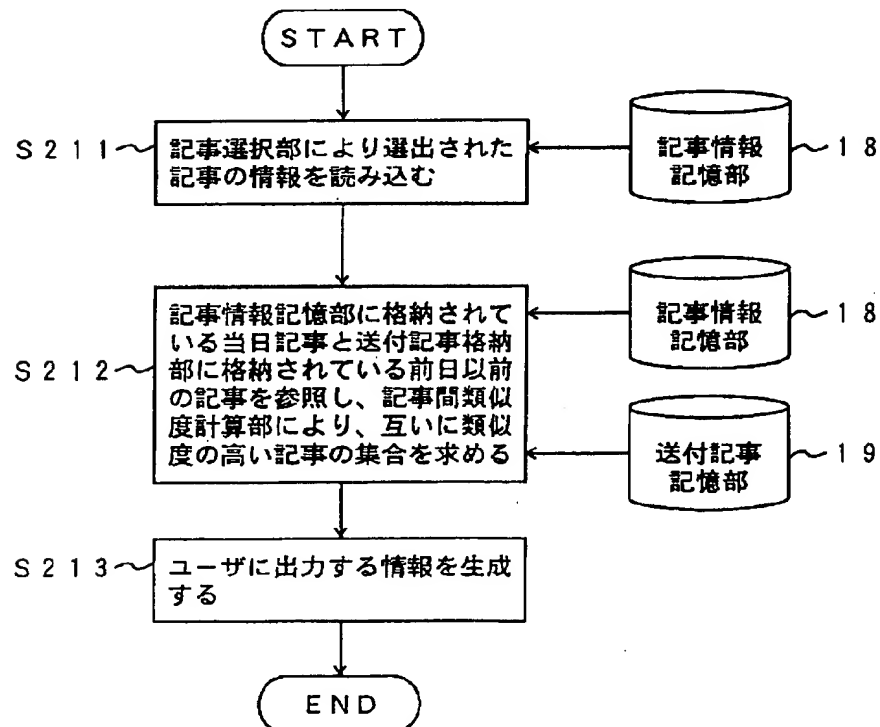
【図87】

1995/08/04
01: マルチメディアの課題電子情報通信学会報告から(8)情報の貴重格差(註)
××産業新聞 p.5 844字
1995/08/03
01: マルチメディアへの課題電子情報通信学会報告から(7)電子図書館
××産業新聞 p.4 883字
1995/08/02
01: マルチメディアへの課題電子情報通信学会報告から(6)ネットワークの料金
××産業新聞 p.5 821字

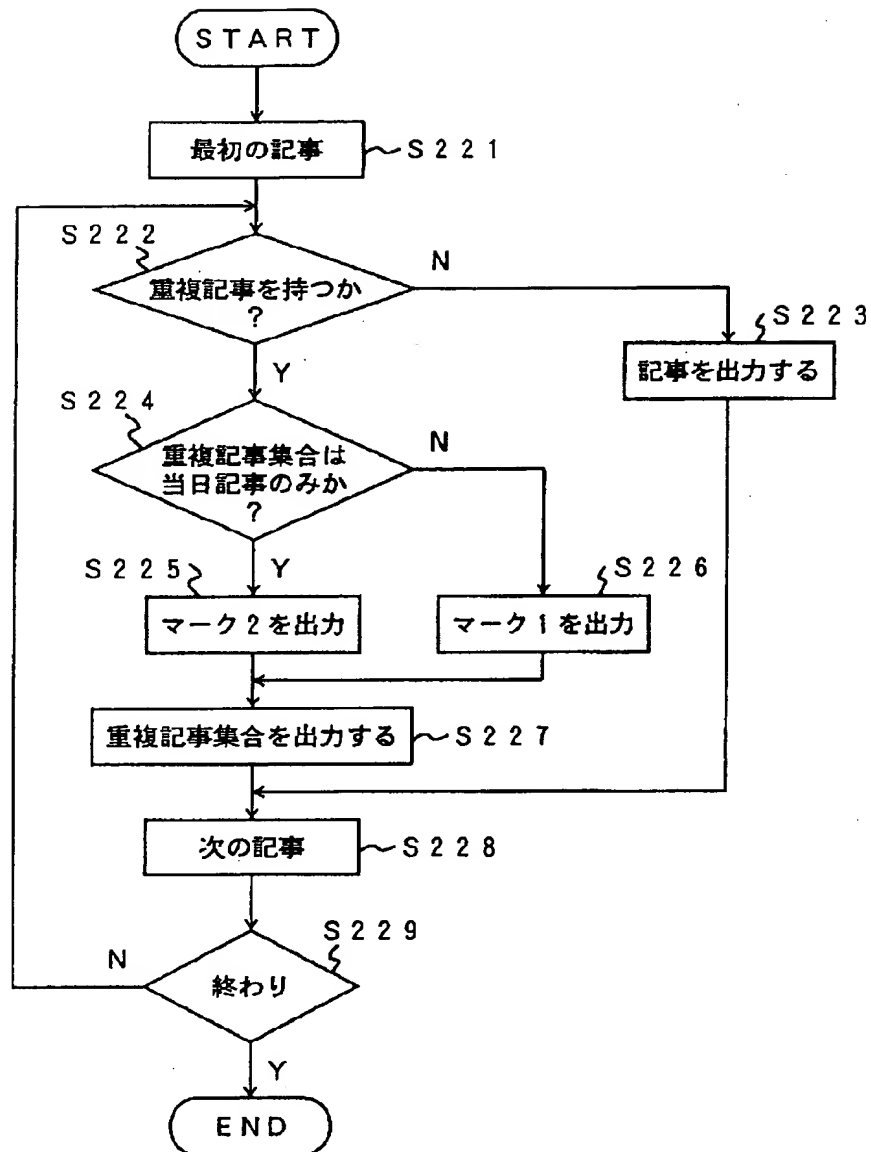
【図80】



【図82】



【図83】



フロントページの続き

(72)発明者 梶浦 正浩
神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

(72)発明者 小野 顕司
神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)